

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic searchr using sw model

Run on: January 31, 2005, 10:52:07 ; Search time 85 Seconds  
(without alignments)  
9482.760 Million cell updates/sec

Title: US-10-048-071-27  
Perfect score: 1134  
Sequence: 1 atgattcaatttcaattaa.....ttaccacagtcagcaaac 1134

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/ptodata/1/ina/5A COMB.seq:  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq:  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:  
5: /cgn2\_6/ptodata/1/ina/PTCUS COMB.seq:  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	849	74.9	3200	3 US-09-381-862-2	Sequence 2, Appli
2	608.2	53.6	1137	4 US-09-583-110-1744	Sequence 1744, Ap
3	608.2	53.6	21338	4 US-08-961-527-20	Sequence 20, Appl
4	317	28.0	1155	4 US-09-107-532A-902	Sequence 502, App
5	247.4	21.8	1158	3 US-09-134-001C-2329	Sequence 2329, Ap
6	236.2	20.8	2347	4 US-08-956-171E-153	Sequence 153, App
7	236.2	20.8	2347	4 US-08-781-986A-153	Sequence 153, App
8	159.6	14.1	546	4 US-09-134-000C-1634	Sequence 1634, Ap
9	59.2	5.2	640681	4 US-09-790-988-1	Sequence 1, Appli
10	57.2	5.0	1141	4 US-09-808-708B-22	Sequence 22, Appl
11	54.8	4.8	1141	4 US-09-806-708B-22	Sequence 22, Appl
12	49.6	4.4	426	4 US-09-134-000C-1635	Sequence 1635, Ap
13	49	4.3	640681	4 US-09-790-988-1	Sequence 1, Appli
14	47.4	4.2	14066	4 US-09-601-198-56	Sequence 56, Appl
15	46.4	4.1	471	4 US-09-543-681A-1232	Sequence 1232, Ap
16	45.8	4.0	751	4 US-08-956-171E-892	Sequence 892, App
17	45.8	4.0	751	4 US-08-781-986A-892	Sequence 892, App
18	45.6	4.0	832	4 US-09-621-976-2813	Sequence 2813, Ap
19	45.4	4.0	821	3 US-08-998-416-541	Sequence 541, App
20	45	4.0	1134	4 US-09-601-198-62	Sequence 62, Appl
21	44.8	4.0	423	4 US-09-710-279-1771	Sequence 1771, Ap
22	44.8	4.0	861	4 US-09-710-279-1145	Sequence 1145, Ap
23	44.8	4.0	984	3 US-09-134-001C-2705	Sequence 2705, Ap
24	44.8	4.0	3315	4 US-09-710-279-3820	Sequence 3820, Ap
25	44.8	4.0	3801	4 US-09-710-279-4271	Sequence 4271, Ap
26	44.8	4.0	3926	4 US-09-710-279-4300	Sequence 4300, Ap
27	43.8	3.9	744	4 US-09-248-796A-778	Sequence 778, App

28	43.4	3.8	15598	4 US-08-956-171E-82	Sequence 82, Appl
29	43.4	3.8	15598	4 US-08-781-986A-82	Sequence 82, Appl
30	43.4	3.8	1664976	4 US-08-916-421B-1	Sequence 1, Appli
31	43.4	3.8	1664976	4 US-09-692-570-1	Sequence 13, Appl
32	42.8	3.8	6113	4 US-10-204-708-13	Sequence 208, App
33	42.4	3.7	3001	4 US-09-539-333D-208	Sequence 6131, Ap
34	42.2	3.7	2919	4 US-09-248-796A-6131	Sequence 5237, Ap
35	41.6	3.7	1782	4 US-09-248-796A-5237	Sequence 11532, A
36	41.6	3.7	2517	4 US-09-893-600-1	Sequence 44, App
37	41.4	3.7	603	4 US-09-248-796A-11532	Sequence 44, App
38	41.4	3.7	825	4 US-09-248-796A-444	Sequence 23, Appl
39	41.4	3.7	1055	4 US-09-806-708B-23	Sequence 167, App
40	41.2	3.6	741	4 US-09-601-198-167	Sequence 3017, Ap
41	41.2	3.6	1347	4 US-09-248-796A-3017	Sequence 3, Appli
42	41	3.6	1425	1 US-07-715-184-3	Sequence 6, Appli
43	41	3.6	1425	1 US-07-876-280-6	Sequence 27, Appl
44	41	3.6	1425	1 US-07-876-280-27	Sequence 1, Appli
45	41	3.6	1425	1 US-07-935-310A-1	

## ALIGNMENTS

RESULT 1  
US-09-381-862-2  
; Sequence 2, Application US/09381862  
; Patent No. 6245906  
; GENERAL INFORMATION:  
; APPLICANT: Ueyama, Hiroshi  
; APPLICANT: Abe, Kanako  
; APPLICANT: Keshi, Hiroyuki  
; APPLICANT: Matsuhisa, Akio  
; TITLE OF INVENTION: PROBES FOR THE DIAGNOSIS OF INFECTIONS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker Drive/6300 Sears Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/381,862  
; FILING DATE:  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; PRIOR APPLICATION NUMBER: JP 1997-71077  
; FILING DATE: 25-MAR-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP98/01288  
; FILING DATE: 23-MAR-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cawley, Jr., Thomas A.  
; REGISTRATION NUMBER: 40,944  
; REFERENCE/DOCKET NUMBER: 19036/36274  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 474-6300  
; TELEFAX: (312) 474-0448  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3200 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus pyogenes

STRAIN: Clinical Isolate SP-7-44  
US-09-381-862-2

Query Match 74.9%; Score 849; DB 3; Length 3200;  
Best Local Similarity 98.7%; Pred. No. 9.5e-204;  
Matches 866; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

```

Qy 1 ATGATTCATTTTCAATTAATGCGCATTTATTATTCATGCTTTTAAATACAACTAAACGT 60
Db 2324 ATGATTCATTTTCAATTAATGCGCATTTATTATTCATGCTTTTAAATGCAACTAAACGT 2383
Qy 61 GCTATTACACTAAAAATGCCATTCCTATTCTTTTCATCAATAAAAAATGGAAGTCACCTCT 120
Db 2384 GCTATTACACTAAAAATGCCATTCCTATTCTTTTCATCAATAAAGATGGAAGTCACCTCT 2443
Qy 121 ACAGAGTAACTTTAAACGGGCTTAACGGTCAATATCAATATGAAAAACACTATTCCTGTA 180
Db 2444 ACAGAGTAACTTTAAACGGGCTTAACGGTCAATATCAATATGAAAAACACTATTCCTGTA 2503
Qy 181 AGTAATGAATGCTGGTTTGTCTAATTTACCTCCAGGAGCTATTTTATTAGAGCTAGT 240
Db 2504 AGTAATGAATGCTGGTTTGTCTAATTTACCTCCAGGAGCTATTTTATTAGAGCTAGT 2563
Qy 241 TTTTATTATTAATTTTCAAGTTTGGCAGATATTAGTATATAAATGTTTAAAGAAATTGAA 300
Db 2564 TTTTATTATTAATTTTCAAGTTTGGCAGATATTAGTATATAAATGTTTAAAGAAATTGAA 2623
Qy 301 CAACACCAAGTTGTTTAAACAGTGGTAAATCAGAGATACCTTAAAGGAAAAGATGTT 360
Db 2624 CAACACCAAGTTGTTTAAACAGTGGTAAATCAGAGATACCTTAAAGGAAAAGATGTT 2683
Qy 361 GACAGTATCCTCGTCAAGAGATATCAAGAGAAATCCTTTCATTTTAAAGCAAAA 420
Db 2684 GACAGTATCCTCGTCAAGAGATATCAAGAGAAATCCTTTCATTTTAAAGCAAAA 2743
Qy 421 TTATTGAAGTCTATTATTGCTGAAACAGCTTTTGCAGCCAGTTTCAAGAAAGTCGCT 480
Db 2744 TTATTGAAGTCTATTATTGCTGAAACAGCTTTTGCAGCCAGTTTCAAGAAAGTCGCT 2803
Qy 481 ATTTTAAACAGGAGTTCATATTGTTATTAAGTATCAATGAATTTTAAAGCAGTAGCGACT 540
Db 2804 ATTTTAAACAGGAGTTCATATTGTTATTAAGTATCAATGAATTTTAAAGCAGTAGCGACT 2863
Qy 541 GACTCTCATCGTATGAGCAACAGTTTAAATCACTTTGGAC-AAATCTTCAGCAGATTTCAT 599
Db 2864 GACTCTCATCGTATGAGCAACAGTTTAAATCACTTTGGACAAATACCTTCAGCAGATTTCAT 2923
Qy 600 GGTAGTCTTCCAAAGTAAATCTTTGAGAGAAATTTTCAGCAGTATTTACAGATGATTTGA 659
Db 2924 TGTGGTTATTCCAAAGTAAATCTTTGAGAGAAATTTTCAGCAGTATTTACAGATGATTTGA 2983
Qy 660 GACCGTTGAGGTATTTTCTCACCAGCCAAATCTTGTTCAGAGTGGAACATTTCTTT 719
Db 2984 GACCGTTGAGGTATTTTCTCACCAGCCAAATCTTGTTCAGAGTGGAACATTTCTTT 3043
Qy 720 TTATACACCGCTCTTAGAAGGAAATATCCCGATACAGACCGCTTTTAAATGACAGAAIT 779
Db 3044 TTATACACCGCTCTTAGAAGGAAATATCCCGATACAGATCGTTTAAATGACAGAAIT 3103
Qy 780 TGAGACGGAGGTGTTTTCATACCAATCCCTTCGCGACGCTATGGAACGTCCTTCTT 839
Db 3104 TGAGACGGAGGTGTTTTCATACCAATCCCTTCGCGACGCTATGGAACGTCCTTCTT 3163
Qy 840 GATTCTTAATGCTACTCAAAATGGTACTGTTAAGCTT 876
Db 3164 GATTCTTAATGCTACTCAAAATGGTACTGTTAAGCTT 3200

```

RESULT 2  
US-09-583-110-1744  
; Sequence 1744; Application US/09583110  
; Patent No. 6699703  
; GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al.  
TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
FILE REFERENCE: PATH00-07A  
CURRENT APPLICATION NUMBER: US/09/583,110  
CURRENT FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/107,433  
PRIOR FILING DATE: 1998-06-30  
PRIOR APPLICATION NUMBER: US 60/085,131  
PRIOR FILING DATE: 1998-05-12  
PRIOR APPLICATION NUMBER: US 60/051,553  
PRIOR FILING DATE: 1997-07-02  
NUMBER OF SEQ ID NOS: 5322  
SEQ ID NO 1744  
LENGTH: 1137  
TYPE: DNA  
ORGANISM: Streptococcus pneumoniae  
US-09-583-110-1744

Query Match 53.6%; Score 608.2; DB 4; Length 1137;

Best Local Similarity 71.1%; Pred. No. 2e-143;

Matches 805; Conservative 0; Mismatches 328; Indels 0; Gaps 0;

```

Qy 1 ATGATTCATTTTCAATTAATGCGCATTTATTATTCATGCTTTTAAATACAACTAAACGT 60
Db 1 ATGATTCATTTTCAATTAATGCGCATTTATTATTCATGCTTTTAAATACAACTAAACGT 60
Qy 61 GCTATTAGCACTAAAAATGCCATTCCTATTCTTTTCATCAATAAAAAATGGAAGTCACCTCT 120
Db 61 GCTATTAGCTTAAAAATGCCATTCCTATTCTTTTCATCAAGTAAAAATTTAGCTGACCAAT 120
Qy 121 ACAGAGTAACCTTTAAACAGGGTCTAACGGTCAAAATATCAATTTGAAAAACACTATTCCTGTA 180
Db 121 GAAGGTATTTTAAATGTTGGTTCAAATGGTCAATTTCAATTTGAAAAATTTTATTTCTCAA 180
Qy 181 AGTAATGAATGCTGGTTTGTCTAATTTACCTCTCCAGAGCTATTTTATTAGAGCTAGT 240
Db 181 AAAAATGAAGTGGTGGTTGTTAAATTAATCTCTTTTAGGTTCGATCTCTTTTGAAGCTCT 240
Qy 241 TTTTATTATTAATTTTCAAGTTTGCAGATATTAGTATATAAATGTTTAAAGAAATTTGAA 300
Db 241 TTTTATTATTAATTTTCAAGTTTGCAGATATTAGTATATAAATGTTTAAAGAAATTTGAA 300
Qy 301 CAACACCAAGTGTGTTTAAACAGTGGTAAATCAGAGATTTACCTTTAAAGGAAAAAGATGTT 360
Db 301 CAAATCAAATGTTTAAACAGTGGTAAATCAGAGATTTACCTTTAAAGGAAAAAGATGAGC 360
Qy 361 GACAGTATCTCTGTACAGAGATATCAACAGAAATCCCTTTGATTTTAAAGCAAAA 420
Db 361 GAACAATATCCCAAGATCCCAAGAAATTTTCAGCAAGCACTCTCTTTAATCTTTGAAACAAA 420
Qy 421 TTATTGAAGTCTATTATTGCTGAAACAGCTTTTGCAGCCAGTTTCAAGAAAGTCGCTCT 480
Db 421 TTACTCAAGAAATTTTAAATGAACAGCTTTTGTGCAAGTACACAGAGAGCTGCTCG 480
Qy 481 ATTTTAAACAGAGTTCATATTGTTATTAAGTAAATCATAAAGATTTTAAAGCAGTAGCGACT 540
Db 481 ATTTTAAACAGGTGCTCACTTCGTATTGAGTCAACACAAAGAGTTTAAAAACAGTTGCAACA 540
Qy 541 GACTCTCATCGTATGAGCCACAGTTTAAATCACTTTGGCAATATCTTCAGCAGATTGATG 600
Db 541 GACTCTCATCGCTTAAGCCAGCAAAAATTTGACTCTTTGAAAAAATAGTGTGATTTTGAT 600
Qy 601 GTAGTTCTTCCAAGTAAATCTTTGAGAGAAATTTTTCAGCAGTATTTTACAGATGATATTGAG 660
Db 601 GTGCTAATCTTCAAGCTTCTCTACGCGAAATTTTTCAGCGGTATTTTACAGATGATATCGAA 660
Qy 661 ACCGTTGAGGTATTTTCTCCAGGCAAAATCTTTGTTTCAAGTGAACATTTCTTTT 720
Db 661 ACTGTAGAGATTTTCTTTTGCCCAATTAACCAAAATCTCTTTTAGAAGCGAAAAATTTAGCTTC 720
Qy 721 TATACACGCTCTTTAGAGAGGAAATTTATCCCGATACAGACCGCTTTTATTAATGACAGAAIT 780

```



Db 721 TATACTCGTCTCTAGAGGAACTATCTCTGATACAGATCGCTTGATTCACACAGACTTT 780  
 Qy 781 GAGACGAGGTTGTTTTCATACCCAAATCCCTTCGCCAGCTATGGAACGTCCTTCTTG 840  
 Db 781 AACACTACTATTACTTTTAAATGGTGAACCTTACGCCAGTCAATGGAGCGTCCCGTCTT 840  
 Qy 841 ATTCTTAATGCTACTCAAAATGGTACTGTTAAGCTTGAGTTGAGTTACTCAAAATCATATTCA 900  
 Db 841 TTATCAAGTGGCAGCTCAAAATGGTACTGTTGAACCTTGAATTTAAGATGGGTTGTTAGC 900  
 Qy 901 GCTCATGTTAACTCACTCAGGTTGGTAAAGTAAACGAGGATTTAGATATTGTTAGTCAG 960  
 Db 901 GCCCATGTTCACTCTCCAGAAGTGGTAAAGTAAACGAGGATTTAGATATTGTTAGTCAG 960  
 Qy 961 TCTGGTAGTGAATTAATCAATCAGCTTCAATCAAACTTCACTTATGAGCTTTTAAAGCT 1020  
 Db 961 ACTGGTGAAGATTGACCAATTAGTTTCAACCCAACTTACTGATTGATTCTCTAAAGCT 1020  
 Qy 1021 ATTAAGAGTGAACAGTAAATTTTCAATTTTATCAATTAATTAACACCAAGTACGACCAAA 1080  
 Db 1021 TTAAATAGGAAAGGTGACCAATTAGCTTTATCTCAGCTGTTGCTCCATTTTACTCTTGTG 1080  
 Qy 1081 CCAGCGCATGAGGAGAAAGTTTATCCAAATTAATTAACACCAAGTACGACCAAA 1133  
 Db 1081 CCAGCAGATACTGACGAGACTTCATGCACTTCATACACCAAGTTCGTACAAA 1133

RESULT 3

US-08-961-527-20  
 ; Sequence 20, Application US/08961527  
 ; Patent No. 6420135  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Charles Kunsch  
 ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
 ; NUMBER OF SEQUENCES: 391  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Human Genome Sciences, Inc.  
 ; STREET: 9410 Key West Avenue  
 ; CITY: Rockville  
 ; STATE: Maryland  
 ; COUNTRY: USA  
 ; ZIP: 20850  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 ; COMPUTER: HP Vectra 486/33  
 ; OPERATING SYSTEM: MSDOS version 6.2  
 ; SOFTWARE: ASCII Text  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/961,527  
 ; FILING DATE:  
 ; CLASSIFICATION: 424  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Brookes, A. Anders  
 ; REGISTRATION NUMBER: 36,373  
 ; REFERENCE/DOCKET NUMBER: PB340P1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (301) 309-8504  
 ; TELEFAX: (301) 309-8512  
 ; INFORMATION FOR SEQ ID NO: 20:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 21338 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; US-08-961-527-20

Query Match 53.6%; Score 608.2; DB 4; Length 21338;  
 Best Local Similarity 71.1%; Pred. No. 5,7e-143;  
 Matches 805; Conservative 0; Mismatches 328; Indels 0; Gaps 0;

Qy 1 ATGATTCAATTTTCAATTAATCGCACATTAATTTATTCATGCTTTTAAATACAACTAAACGT 60  
 Db 7076 ATGATTCAATTTTCAATTAATTAATAAATTTATTTCTACAAGCATTAATACTACTAAGAGA 7135  
 Qy 61 GCTATTAGCAGTAAATGCGCAATTCCTATTCTTTTCATCAATAAATAATGAAGTCACTTCT 120  
 Db 7136 GCTATTAGTCTTAAATAATGCCAATTCCTATTATCAACAGTAAATAATGACGTGACCAAT 7195  
 Qy 121 ACAGGAGTAACCTTTAACAGGGTCTAACGGTCAAATATCAATTTGAAAAACAATATTCCTGTA 180  
 Db 7196 GAAGGTATTACTTTAATTTGGTTCAATATGTCATAATTTCAATTTGAAAAATTTTATTTCTCA 7255  
 Qy 181 AGTAATGAATGCTGGTTTGGTAAATTAATCTCTCCAGGAGCTATTTTATTAAGAGCTAGT 240  
 Db 7256 AAAAATGAAGATGCTGGTTTGGTAAATTAATCTCTCTTTAGGTTCGATCCTCTTGAAGCTTCT 7315  
 Qy 241 TTTTATTAATTAATTTTCAAGTTTGGCAGATATTTAGTATATAATGTTTAAAGAAATTTGAA 300  
 Db 7316 TTTTATTAATTAATTTTCAAGTTTGGCAGATATTTAGTATATAATGTTTAAAGAAATTTGAA 7375  
 Qy 301 CAACACCAAGTTTGTTTTAAACAGTGTAAATCAGAGATTTACCTTAAAGGAAAAAGATGTT 360  
 Db 7376 CAAATCAAAATTTGTTTAAACAGTGTAAATCAGAGATTTACCTTAAAGGAAAAAGATGTT 7435  
 Qy 361 GACAGTATCTCGTCTCAAGAGATATCAACAGAAATTCCTTTGATTTTAAAGAAACAAA 420  
 Db 7436 GAACAATATCCACGAATCCAGAAATTTTCAGCAAGCACTCTCTTAAATTAATCTTCAAGAA 7495  
 Qy 421 TTTTGAAGTCTATTATTGCTGAAACAGCTTTTGCAGCAGCTTTTCAAGAAAGTCTGCT 480  
 Db 7496 TTACTCAAGAAATTTAATTAAGAAACAGCTTTTGCAGCAGCTTTTCAAGAAAGTCTGCTG 7555  
 Qy 481 ATTTTAAACAGGAGTTTCAATTTGTTAATTAAGTAAATCAATAAAGATTTTAAAGCAGTAGCGACT 540  
 Db 7556 ATTTTAAACAGGAGTTTCAATTTGTTAATTAAGTAAATCAATAAAGATTTTAAAGCAGTAGCGACT 7615  
 Qy 541 GACTCTCATGTAAGAGCAAGCTTTAATCACTTTTGGCAATATCTTTCAGCAGATTTGATG 600  
 Db 7616 GACTCTCATGCTTAAAGCCAGAAAAATTTGACTCTTTGAAAAAATAGTATGATTTGAT 7675  
 Qy 601 GTAGTTCTTCCAAAGTAAATCTTTGAGAGAAATTTTTCAGCAGATTTTACAGATGATATTGAG 660  
 Db 7676 GTGGAATTTCTAGCGCTTCTTACCGAAATTTTTCAGCGGATTTTACAGATGATATTGAG 7735  
 Qy 661 ACCGTTGAGGTATTTTCTCACCAGCCAAATCTTTGTTTCAAGAGTGAACACATTTCTTTT 720  
 Db 7736 ACTGTAGAGATTTTCTTTGCCAAATTAACCAATCTCTTTAGAGCGGAAATATTAGCTTC 7795  
 Qy 721 TATACACGCTCTTAGAAGGAAATTTATCCGATACAGACCGTTTATTAATGACAGAAATTT 780  
 Db 7796 TATACGCTCTCTTAGAAGGAAATTTATCCGATACAGATCGCTTGTGATTTCCAAACAGACTTT 7855  
 Qy 781 GAGACGAGGTTGTTTTCATACCAATCCCTTCGCCAGCTATGGAAGCTGCTCTTCTTG 840  
 Db 7856 AACACTACTATTACTTTAATTTGTTGTAACCTTACCCAGTCAATGGAGGTCGCCGCTTT 7915  
 Qy 841 ATTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTTGAGATTACTTCAAAATCATATTTCA 900  
 Db 7916 TTAACAAGTGGAGTCAAAATGGTACTGTTGAAATTTGAAATTAAGATGGGGTTGTAGC 7975  
 Qy 901 GCTCATGTTAACTACCTCAGGTTGGTAAAGGTAAGGTAAGGATTTTATGATTTAGTACG 960  
 Db 7976 GCCCATGTTCACTCTCCAGAAAGTTGGTAAAGTAAACGAAGAAATCGATACATGATCAGGTT 8035  
 Qy 961 TCTGGTAGTGAATTTAATCACTATCAGCTTCAATCCAACTTACCTTATTTAGTCTTTAAAGCT 1020  
 Db 8036 ACTGGTGAAGATTTGACCAATTAGTTTCAACCCAACTTACTTGAATTTCTCTTAAAGCT 8095  
 Qy 1021 ATTTAAAGTGAAACAGTAAATTTCAATTTCTTATCACCAGTTTCGACCAATTCACCCCTAACA 1080  
 Db 8096 TTAATACGAAAGGAGTCACTATTAGCTTTATCTCAGCTGTTGCTGCAATTTACTCTTGTG 8155  
 Qy 1081 CCAGCGATGAGGAGAAAGTTTATCCAAATTAATTAACACCAAGTACGAAACAAA 1133

Db 8156 CCAGCAGACTACTGACGAAGACTTCATGCGAGCTCAATACACGAGTTCGTACAAA 8208

RESULT 4  
US-09-107-532A-902  
; Sequence 902, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arinello, Pamela Deneka  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 902:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1155 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (B) LOCATION 1...1155  
; SEQUENCE DESCRIPTION: SEQ ID NO: 902:  
US-09-107-532A-902

Query Match 28.0%; Score 317; DB 4; Length 1155;  
Best Local Similarity 55.6%; Pred. No. 2.7e-70;  
Matches 630; Conservative 0; Mismatches 500; Indels 3; Gaps 1;

Qy 1 ATGATTCAATTTTCAATTAATCGACATTTATTTATTCATGCTTTTAAATACAACTAAACGTT 60  
Db 22 ATCATGAAAGTTTACTTTAAACCGAGTAGCTTTATGCGAGGATTCGAACTGTTTCAACGA 81

Qy 61 GCTATTAGCACTAAAAATGCCATTCCTATTCCTTCATCAATTAATAATGAAGTCACTTCT 120  
Db 82 GCTATTTTCAAGCAAAACCAACGATCCCTATTTTTCAGAGGTGTAAAAATCACAATGACA 141

Qy 121 ACAGAGTAACCTTTTAAACGGGTCTAACGCTCAAAATCAATTAATGAAGAAACACTATTCCTGTA 180  
Db 142 GAAGGTTTGACTTTGACGGGGAGCAACGCTGATATATCAATTAATGAAGAACTTTTGTCTGTT 201

Qy 181 AGTAATCAAAATGCTGTTTCTTAATTAATCTCTCCAGGAGCTATTTTATTAGAAGCTAGT 240  
Db 202 GAAAAACGAAAAAGCAAATATGCAATATGCAATCTACTGTTCCATTTGTTTACAAGCAGCT 261

Qy 241 TTTTATTATATATATTTCAAGTTTCCAGATATTAGTATATATATATATATATATATATAT 300  
Db 262 TTCTTTAGCGAAATCAATTCGGAGACTTCTCTGGAAGAAACATTTACTTTTGAAGTTTGTAG 321

Qy 301 CAACACCAAGTTGTTTAAACCAAGTGTAAATTCAGAGATTACCTTTAAAGGAAAAAGATGTT 360  
Db 322 AATPAAACAGTAGGATGATCACTTCTGGAAGAGGAAATTTATCGTAATGGATTAGATGCA 381

Qy 361 GACCAGTATCTCTCTACAGAAAGATATCAACAGAAATCTCTTTGATTTTAAACAAAA 420  
Db 382 GATAACTATCT 441

Qy 421 TTATTGAAGTCTATTATTGCTGTAACAGCTTTTTCAGCCAGTTTACAGAAAGTCGCTCT 480  
Db 442 GTATTGACTAACTAATCAACAGAAACAGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 501

Qy 481 ATTTTAAACAGGAGTTTCATATTGTAATTAAGTAATCATAAAGATTTTAAAGCAGTAGCCACT 540  
Db 502 ATCTTGACAGGTGTCATTTTCAATTTA---TCGTATAATCTTTTATTAGCTGTAGTACT 558

Qy 541 GACTCTCATGATGAGCCAAAGTTTAAATCACTTTGGACAATATCTTCAGCAGATTGATG 600  
Db 559 GATTCTCACGCTTAAAGTCAACGCGTATTCAGTAGAAACAAAGCGGTGATCATTTTGAT 618

Qy 601 GTAGTTCTTCCAGTAAATCTTTGAGAGAAATTTTTCAGCAGTATTTACAGATGATATGAG 660  
Db 619 ATTGTTATCTGGAAGAAAGTTTGTATCGAATTTATCTGCTCATTAACAAATGAAGAANA 678

Qy 661 ACCGTGAGGTATTTTCTCACCAAGCCAAATCTTGTTCAGAAAGTGAACACATTTCTTTT 720  
Db 679 ATCGTCGAATCAGCATTAATGAGAAACCAAGTGTGTTTCAAAACAGAAACGATGATTTTC 738

Qy 721 TATACAGCCTCTTTAGAGGAAATATCCGATACAGACGCTTTTATTAATGACAGAAATTT 780  
Db 739 TATTCCGTTTGTCTAGAGGAAACATATCCAGATACCAATCGTTTGTGATTTCTTCAAGTTT 798

Qy 781 GAGACGGAGGTGTTTTCATACCCATCCCTTCGCCACGCTATGGAACGTCCTTCTTG 840  
Db 799 AATACAGAGGTGAAATTTCTGTCCTAGCTTTTTCAGGCGATCGAACGGGCTTCTTTA 858

Qy 841 ATTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTGAGATTACTCAAAATCATATTTCA 900  
Db 859 CTTTCTCATGAAGCGCGTAACAACATCGTTCTGTTTATCTATTCGTCAGATGCGGTTGTT 918

Qy 901 GCTCATGTTAACTCACCTGAGGTGGTAAGTAACGAGGATTTAGATATTGTTAGTCAG 960  
Db 919 TTATACGGAATTCGCCAGAAATCGGAAAGTAGAAGAAAGCTTGAGTTATACAGCAAGC 978

Qy 961 TCTGTAGTATTAACTATCAGCTTCAATCCAACTTACCTTATTGAGTCTTTTAAAGCT 1020  
Db 979 AGTGGGATCCATAGATATTTCTTCAATCCGGAATTAATGAAGACGCTTGGGTGCA 1038

Qy 1021 ATTAAGTGAACAGATAAAATTCATTTCTTATCACAGTTTCGACCAATTCACCTTAACA 1080  
Db 1039 TTTGAGATATGACATCAAAAGTGAATTTATCTCTGCTATTCGTCATTTTACATTGGA 1098

Qy 1081 CCAGCGGATAGGAGAAAGTTTATCCAAATTAATTAACAGCAGTACGACGAA 1133  
Db 1099 CCAACGGAAGATGGCGTTTCAATTTTATTCAGCTGATTACTCTGTACGTACAAA 1151

RESULT 5  
US-09-134-001C-2329  
; Sequence 2329, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007  
 CURRENT APPLICATION NUMBER: US/09/134,001C  
 CURRENT FILING DATE: 1998-08-13  
 PRIOR APPLICATION NUMBER: US 60/064,964  
 FILING DATE: 1997-11-08  
 PRIOR APPLICATION NUMBER: US 60/055,779  
 FILING DATE: 1997-08-14  
 NUMBER OF SEQ ID NOS: 5674  
 SEQ ID NO 2329  
 LENGTH: 1158  
 TYPE: DNA  
 ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-2329

Query Match 21.8%; Score 247.4; DB 3; Length 1158;  
 Best Local Similarity 52.9%; Pred. No. 8.1e-53;  
 Matches 580; Conservative 0; Mismatches 511; Indels 6; Gaps 2;

QY 1 ATGATTCATTTTCAATTAATCGCACATATTTTATTCATGCTTTTAAATACACATTAACGCT 60  
 DB 25 ATGATGGAATTCACAAATTAAGAGAGATTATTTTATTAATCACTTTAACGACACATTAAG 84  
 QY 61 GCTATTAGCACTAAAAATGCCATTCCTATTCTTTCATCAATAAAAAATTTGAAGTCACTTCT 120  
 DB 85 GCTATCTCACCTAGAACAACTTTACGATTTTAAACAGGTATCAAAATTTGATGCTTAAGAA 144  
 QY 121 ACAGAGTAACCTTTAACAGGGTCTAACGGTCAAAATATCAATTTGAAAAACATTTCC--T 177  
 DB 145 AATGAAGTCATTTTAACTGGGTGCAGATTCAGAAATATCAATAGAAATCACTATTCCAAAA 204  
 QY 178 GTAAGTAATGAAGTCTGGTTGGTAAATTTACCTCTCAGGAGCTATTTTATTAGAAGCT 237  
 DB 205 CAAGTTGATGGTGAAGAAATTTGTAATTTACAGAAACAGGATCAGTTGTACTTCCCTGGT 264  
 QY 238 AGTTTTTTTATTAATATTTTCAAGTTTGCAGATATTAGTATATAAATGTTAAAGAAATTT 297  
 DB 265 CGTTTCTCGTTGATATTTAATAAACTCTCGAAAGAAAGTTAAATTTATCACTAAT 324  
 QY 298 GAACAACACCAAGTTGTTTAAACAGTGGTAAATCAGAGATTACCTTTAAAGGAAAAAGAT 357  
 DB 325 GAACAATTTCCAAACGTTAATACATCAGCTCATTCAGAAATTTAACTTAAGTGGCTTAGAT 384  
 QY 358 GTTGACAGTATCCTCGTCTACAGAAAGTATCAACAGAAATTCCTTTGATTTTAAACA 417  
 DB 385 CCTGATCAATACCCATTTATCCTGAGGTATCAGAGATGACGCTATTCAATTTGTCAGTT 444  
 QY 418 AAATTTTGAAGTCTATTATTCTGCTCAACAGCTTTTGCAGCCAGTTTCAAGAAAGTCTGT 477  
 DB 445 AAGTACTAATAAATATCATGCAAACTAATTTTGCAGTGTCCACCTCAGAAACAGCA 504  
 QY 478 CTTATTTTAAACGAGGTTCATATTGTTAAGTAAATCAATAAAGATTTTAAAGCAGTAGCG 537  
 DB 505 CCAGTACTTACTGGTAACTGGCTTATA---CAAGATAATGAATTAATATGCACAGCA 561  
 QY 538 ACTGACTCTCATCGTATGAGCCAAAGTTTAACTTTGGAACAATCTTACGACGATTTG 597  
 DB 562 ACAGATTCACACCGCTTAGCTGTAAGAAAGTTTACAGTTTGAAGATGCAATCAGAAATAAA 621  
 QY 598 ATGGTAGTTCTTCCAAAGTAAATCTTTTGAGAGAAATTTTTCAGAGATTTTACAGATGATTT 657  
 DB 622 AATGTCATATTCCTGGTAAAGCTTTTCTGAAATTAACAAATTAATGAGTGACAGGAC 681  
 QY 658 GAGACCGTTGAGGTATTTTCTTCCAAAGCCAAATCTTTGTTTCAAGAGTGAACACATTTCT 717  
 DB 682 GAAGATATTTGATATTTCTTTCTTCTTAAACCAAGTGTATTTCAGAGTGGGAATTAAT 741  
 QY 718 TTTTATACAGCGCTCTTGAAGAAATTTATCCGATACAGACCGTTTATTATATGACAGAA 777  
 DB 742 TTTCTCTCAGCTTTTAACTGAGGTCAATTTATCCAGATACGACAGCTTTTATTTCCAGAAAT 801  
 QY 778 TTTGAGACGAGGTGTTTCTTCAATACCCATCCCTTCCGACGCTATGGAAGTCCTTC 837  
 DB 802 TATGAGATTAATTAAGAAATTAACAAATGAGAGACTTCTATCATGCAATTTGATCGTCACT 861

QY 838 TTGATTTCTAATGCTACTCAAAATGGTACTGTTTAAGCTTGAGATTACTCAAAATCATATTT 897  
 DB 862 TTATTTAGCACGTGAAGGTGGAATAATGTTTATTAATTAAGTACAGGTAATGAATTAGTT 921  
 QY 898 TCAGTCTATGTAATCACTCACCTGAGGTGGTAAAGTAAACGAGGATTTAGATATTTAGT 957  
 DB 922 GAATCTTTCATCTACTTCTCCTGAAATTTGTTACTGTTTAAAGAAAGATTTAACGCTAATGAT 981  
 QY 958 CAGTCTGGTAGTGTATTAATCACTTATCAGCTTCAACTCAACTTACCTTTATGAGTCTTTAAAA 1017  
 DB 982 GTAGAACGCGGAACCTTGAATAATTTCTTTCAACTCAAAATACATGATGATGCTTTAAAA 1041  
 QY 1018 GCTATTTAAAGTGAACAGCAATTAATTTCACTTTCTTATCAGGTTGCGACCATTTCCCTTA 1077  
 DB 1042 GCCATTGATAATGATGAAGTAGAATGATTTCTTTGTTGTAACATGAAACCATTTTCTTTA 1101  
 QY 1078 ACACGAGCGGATGAGGA 1094  
 DB 1102 AAACCAAAAGATGATGA 1118

RESULT 6

US-08-956-171E-153  
 ; Sequence 153, Application US/08956171E  
 ; Patent No. 6593114  
 ; GENERAL INFORMATION:

APPLICANT: Charles Kunsch  
 Gil H. Choi  
 Patrick S. Dillon  
 Craig A. Rosen  
 Steven C. Barash  
 Michael R. Fannon  
 CITY: Rockville  
 STATE: Maryland  
 COUNTRY: USA  
 ZIP: 20850  
 ADDRESS: Human Genome Sciences, Inc.  
 STREET: 9410 Key West Avenue  
 CITY: Rockville  
 STATE: Maryland  
 COUNTRY: USA  
 ZIP: 20850

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
 COMPUTER: HP Vectra 486/33  
 OPERATING SYSTEM: MSDOS version 6.2  
 SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/956,171E  
 FILING DATE: 20-Oct-1997  
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/009,861  
 FILING DATE: January 5, 1996  
 APPLICATION NUMBER: 08/781,986  
 FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION:  
 NAME: Mark J. Hyman  
 REGISTRATION NUMBER: 46,789  
 REFERENCE/DOCKET NUMBER: PB248P1

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (240) 314-1224  
 TELEFAX: (301) 309-8439

INFORMATION FOR SEQ ID NO: 153:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2347 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 153:  
 US-08-956-171E-153

Query Match

20.8%; Score 236.2; DB 4; Length 2347;

		Beet Local Similarity	52.2%; Pred. No. 6.8e-50;		
		Matches	573; Conservative	0; Mismatches	518; Indels
		Gaps	2;		
Qy	1	ATGATTCAAAATTTTCAAATTAATCGCACATAATTATTCATGTCTTTAAAAACAACTAAAAGC	60		
Dd	393	ATGATGGGAATTCACATATTATAAAGAGATTATTTTTATACACAATTAATGACACACTTAAAA	452		
Qy	61	GCTATTAGCATAAAATGCCATTCCTTATTCCTTCATCAATAAAAAATTGAAGTCATTCT	120		
Dd	453	GCTATTTCCACAAGAACAACATTAACCCTATATAATCTGGTATCAAAATCGATSCGAAAGAA	512		
Qy	121	ACAGGAGTAACCTTTAAACAGGCCTAAACGGTCAAAATATCAAAATGAAAAACACTATTCCTGTA	180		
Dd	513	CATGAAGTTATATTAACCTGGTTCAGACTTGAAATTCATAGAAATCATCTATTCCTTAAA	572		
Qy	181	AGT---AATGAAAATGCTGGTTTGCTTAATTAACCTCTCAGAGCTATTTTATTAGAAGCT	237		
Dd	573	ACTGTAGTAGCGAAGATAATTTGCTCAATATTTTCAGAAAACAGGCTCAGTAGTACTTCTCGGA	632		
Qy	238	AGTTTTTTTTTAATAATATTTATTCAGTGTTCACGATATTAGTATAAATGTTAAAGAAAT	297		
Dd	633	CGATCTTTGTGTGATATATAAAAAAATTAACCTGTAAGATGTTAAATATATCTACAAAT	692		
Qy	298	GAACAAACCAAGTTGTTTTAACAGTGGTAAATCAGAGATTACCTTAAAGAGAAAAGAT	357		
Dd	693	GAACAAATCCAGACATAAATFACATCAGGTCAATTCGAAATTAATTTAAGTGGCTTAGAT	752		
Qy	358	GTTACACAGTATCCTCGTCTACAGAAGTATCAACAGAAAATCCTTTGATTTTAAAAACA	417		
Dd	753	CCAGTCAATATCCTTTATTAACCTCCAAGTTTCTAGATGACGCAATTCATTTGTCGGTA	812		
Qy	418	AAATATTGAAGTCTATTTATGCTGAAACAGCTTTTTCAGCCAGCTTTTACAGAAAGTCGT	477		
Dd	813	AAGTGCTTAAAAACGGTATTCACAACAAATTTTCGAGTGTCCACTCAGAAACAGC	872		
Qy	478	CTATTTTAAACAGAGTTCATTTGTTATTAAGTAATCATAAAGATTTTAAAGCAGTAGCG	537		
Dd	873	CCAGTACTAATCTGGTGTGAATCGGCTTATA--CAAGAAAATGAAATTAATATGCACAGC	929		
Qy	538	ACTGACTCTCATCTGATGAGCAACGTTTAAATCACTTTGGACAATCTTCAGACAGATTG	597		
Dd	930	ACTGACTCACACGCTTGGCTGTAAGAAAGTTGCAGTTAGAGATGTTCTGAAAAACAA	989		
Qy	598	ATGTTAGTTCTTCCAGTAATCTTTGAGAGAAATTTTTCAGCAGTATTTACAGATGATATT	657		
Dd	990	AATGTCATCATTCAGGTAAGGCTTTAGCTGAATTAATAAATAAATATGCTGACAAATGAA	1049		
Qy	658	GAGACCGTCGAGTATTTTCTCACAGGCCAAATCTTGTTTCAGAAGTGAACACATTTCT	717		
Dd	1050	GAAGACAATTGATATCTTCTTTGCTTCAACCCAGTTTATTTAAAGTTGGAATGTGAAC	1109		
Qy	718	TFTTTATACGCGCTTTAGAAGGAAATATTCCCGATACAGACCGTTTATTAAATGACAGAA	777		
Dd	1110	TTTTATTCTCGATTATTAGAAGGACATATCTCGATACACACGTTTATTCCTCGAAAAC	1169		
Qy	778	TTTTCAGAGCGGAGTTGTTTTCAATACCCATCCCCTCGCACGCTATGGAACGTGCGCTTC	837		
Dd	1170	TATGAAATTAATTAAGTATAGACAATGGGAGTTTTATCATCGATTGATCGTGGCTCT	1229		
Qy	838	TTGATTTCTAATGCTACTCAAAAATGGTACTGTTAAGCTTCAGATTACTCAAAATCATATT	897		
Dd	1230	TTATTAGCGGTGAAGGTGGTAATAACGTTATTATAATTAGTACAGGTGATGAGCTGTT	1289		
Qy	898	TCAGCTCATGTTAACTCACCTGAGGTGGTAAAGTAAACGAGGATTTAGATATTTGTTAGT	957		
Dd	1290	GAATTGTTCTCATATCACCAGAAATTTGTTACTGTATAAAGAAAGATTCGATGCAAAACGAT	1349		
Qy	958	CAGTCTGGTAGTGAATTTAACTATCAGCTTCAATCCAACTTACCTTATTTAGTCTTTAAAA	1017		
Dd	1350	GTTGAAGGTGGTAGCCTGAAAATTTCAATTCATCTCTAATAATATGATGGATGCTTTAAAA	1409		
Qy	1018	GCTATTAAGAGTGAAACAGTAAAAATTCATTTCTTATCACAGTTCCACATTCACCCCTA	1077		

```

Db      1410  GCAATCGATGAATGATGAGGTTGAAGTTGAAATCTTCGGTCAATGAAACCAATTTATTCTA 146S
Qy      1078  ACACGAGCGCATGAGGA 1094
          |||||  |||||
Db      1470  AAACCAAAAGGTGACGA 1486

RESULT 7
US-08-781-986A-153
; Sequence 153, Application US/08781986A
; Patent No. 6737248
; GENERAL INFORMATION:
;   APPLICANT: Charles Kunsch
;   TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
;   NUMBER OF SEQUENCES: 5255
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Human Genome Sciences, Inc.
;   STREET: 9410 Key West Avenue
;   CITY: Rockville
;   STATE: Maryland
;   COUNTRY: USA
;   ZIP: 20850
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;   COMPUTER: HP Vectra 486/33
;   OPERATING SYSTEM: MSDOS version 6.2
;   SOFTWARE: ASCII Text
;   CURRENT APPLICATION DATA:
;   FILING DATE:
;   APPLICATION NUMBER: US/08/781,986A
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Benson, Bob
;   REGISTRATION NUMBER: 30,446
;   REFERENCE/DOCKET NUMBER: PB248PP
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (301) 309-8504
;   TELEFAX: (301) 309-8512
;   INFORMATION FOR SEQ ID NO: 153:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2347 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
; US-08-781-986A-153

```

Query Match	20.8%; Score 236.2; DB 4; Length 2347;
Best Local Similarity	52.2%; Pred. No. 6.8e-50;
Matches	573; Conservative 0; Mismatches 518; Indels 6; Gaps 2;
Qy	1 ATGATTCAAATTTCAATTAATCGCACATTAATTTATTCATGCTTTAAATACAACCTAAACGT 60
Db	393 ATGATGGAATTCATTTAAAGAGATTAATTTATTACACAATTAATGAACATTAAAA 452
Qy	61 GCTATTAGCACTAAAATGCAATTCCTATTCTTTTCATCAATAAAAAATGGAATCACTTCT 120
Db	453 GCTATTTTCACACAGAACCAATTCACCTATTATTAACCTGGTATCAAAATCGGATGCGAAGAA 512
Qy	121 ACAGGAGTAACCTTTAACACAGGGTCTAACGGTCAAAATATCAATTCGAAAACACTATTTCCTGTA 180
Db	513 CATGAAGTTATATTAACTGGTTCAGACTCTGAATTTCAATAGAAATCACTATTTCCTAAA 572
Qy	181 AGT---AATGAAAATGCTGGTTTGCTAAATACCTCCAGAGGCTATTTTATTAGAACGT 237
Db	573 ACTGTAGATGCGAAGATAATTGTCAATATTTACAGAAACAGGCTCAGTAGTACTTCCTGGA 632
Qy	238 AGTTTTTTTATTAATATATTATTCAGTTTGCCAGATATTAGTATAAATGTTAAAGAAATT 297
Db	633 CGATTCTTTGTCATATATAAAAAAAATTAACCTGGTAAAGATGTTAAATTAATCTACAAAT 692

```

QY 298 GAACAACACCAAGTTGTTTAAACAGTGTGTAATCAGAGATTACCTTAAAGGAAAGAT 357
Db 693 GAACAATCCAGACATAATTAATCATCAGGTCTATCTGAATTTAAATTAAGTGGCTTAGAT 752
QY 358 GTTGACAGTATCTCTCGTCTACAAGAGTATCAACAGAGAAATCCCTTTGATTTTAAACA 417
Db 753 CAGATCAATATCCCTTATTAATCTCAAGTTCTAGAGATGACGAATTCOAATTCGCGTA 812
QY 418 AATATTGAAGTCTATTATGCTGAAACAGCTTTTGCAGCCAGTTTACAAGAAAGTCGT 477
Db 813 AAGTGTCTTAAACAGTATGACACAAACAAATTTTGCAGTGTCCACCTCAGAAACACGC 872
QY 478 CTAATTTAAACAGGATTCATATTGTTAAGTAATCAATAAGATTTTAAACAGTAGCG 537
Db 873 CAGTACTAACTGGGTGGAATGGCTTATA---CAAGAAATGAATTAATATGSCACGC 929
QY 538 ACTGACTCTCATCTGATGAGCAACGTTTAAATCACTTTGCAACAATCTTTCAGCAGATTG 597
Db 930 ACTGACTCACCGCTTGCTGTAGAAAGTTGCGAGTTAGAGATGTTTCTGAAACAAA 989
QY 598 ATGGTAGTTCTTCCAAGTAAATCTTTGAGAGAAATTTTTCAGCAGTATTACAGATGATTT 657
Db 990 AATGTCATCTTCCAGGTAAGCTTTAGCTGAATTAATAATAAATATATGCTCAGCAATGAA 1049
QY 658 GAGACGTTGAGGTATTTTCTCAACAGCCAAATCTTGTTCAGAGTGAACACATTTCT 717
Db 1050 GAAGACATTTGATATCTCTCTGCTTCAACCAAGTTTATTATTAAGTTTGAATGTGAAC 1109
QY 718 TTTTATACAGCTCTTGAAGAAATATCCGATACAGACGTTTATTATATGACAGAA 777
Db 1110 TTTATTTCTCGATTTATGAGAGACATATCTCTGATACACACGTTTATCTCCCTGAAAC 1169
QY 778 TTTGAGACGAGGTTGTTTCAATACCAATCCCTTTGCGCAGCTATGGAACGTCCTTTC 837
Db 1170 TATGAAATTTAAATTAAGTATACAAATGGGAGTTTATCATGCGATGATGTCGCTCT 1229
QY 838 TTGATTTCTAATGCTACTCAAAATGCTAGTTTAAAGTTGAGATTAATCAAAATCATATT 897
Db 1230 TTTATGCGCGTGAAGGTGTAATACGTTTATTAATTAAGTACAGGTGATGACGCTGTT 1289
QY 898 TCAGCTCATGTTAACTCACTCAGGTGTTGTAAGTAAAGGATTAAGATTTAGATATTGTTAGT 957
Db 1290 GAATTTGCTTCTACATCACAGAAATGTTACTGTAAAGAAAGAGTGTGATGCAAAACGAT 1349
QY 958 CAGTCTGTTAGTATTAACTATCAGCTTCAATCAACTTACCTTATTGAGTCTTTTAAAA 1017
Db 1350 GTTGAAGGTGTTAGCTGAAATTTCAATCACTCTAAATATATGATGATGATGCTTTAAA 1409
QY 1018 GCTATTAAAGTGAACAGTAAATTCATTTCTATACCAAGTTCGACCATTCACCTTA 1077
Db 1410 GCAATCGATAATGATGAGGTGAAGTTGAATTTCTTCGGTACAAATGAAACCAATTTATTCTA 1469
QY 1078 ACACCAAGGATGAGGA 1094
Db 1470 AAACCAAGGTTGACGA 1486

```

```

RESULT 8
US-09-134-000C-1634
; Sequence 1634, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1634

```

```

; LENGTH: 546
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-1634
Query Match 14.1%; Score 159.6; DB 4; Length 546;
Best Local Similarity 55.9%; Pred. No. 6.9e-31;
Matches 303; Conservative 0; Mismatches 239; Indels 0; Gaps 0;
QY 592 GATTTGATGGTAGTTCTTCCAAAGTAAATCTTTGAGAGAAATTTTTCAGCAGTATTTCAGAT 651
Db 1 GACTTTAATGTTGTAATTTCCAGGAAAGTAAATTTTCTGTTTCTGTTTCAATTAACCAAT 60
QY 652 GATATTGAGACCGTTGAGGTATTTTCTCACCAGCCAAATCTTGTTCAGAAAGTGAACAC 711
Db 61 GAAGAAGAAATGGTTGAAATCAGCATTTATGGAATAATCAAGTCTATTATTAACACAGAAACA 120
QY 712 ATTTCTTTTATACAGGCTCTTAGAAGAAATTTATCCGATACAGACGTTTATTAATG 771
Db 121 ATGTACTTTCTATTCTCGTTTGTAGAGGAAATTTATCTGTATACCAACCGTTTAAATTTCCA 180
QY 772 ACAGAAATTTGAGACGAGGTTGTTTCAATACCCAAATCCCTTCGCCACGCTATGGAACGT 831
Db 181 ACTAGCCATAACACCAATTTGAATTTATGTACAGAAATGCTTTTCAGCAATCGAAGCT 240
QY 832 GCCTTTCTGATTTCTTAATGCTACTCAAAATGGTACTGTTAAAGCTTGAATTAATCTCAAAAT 891
Db 241 GCCTTTCTTCTCATGAAGGAGCTAAACAATATTGTTTCGCTTTCAATTTTCCAGAT 300
QY 892 CATATTTCAGCTCATGTTTAACTCAGCTGAGGTGTAAGTAAAGTAAAGAGATTTAGATATT 951
Db 301 TCTGTTGTTTATATGGAATTTTCACTGAAATTTGGAAGTTCGAAAGAGCTTTTAAACTAT 360
QY 952 GTTACTCAGTCTGTTGAGTATTAACTATCAGCTTCAATCCAACTTACCTTATTCAGTCT 1011
Db 361 GAAATGTTTCTGTTGAGGCGTTGGATATTCTTTCAACCCAGATTAATGAAGATGCG 420
QY 1012 TTAAGATTTAAAAAGTGAACAGTAAAAATTTCAATTTCTTATCACCAGTTCGACCATTC 1071
Db 421 TTGCGGCGTTTGGCGATATGAATATTACCGTGAAATTTCTTCTCAATTCGTCGTTT 480
QY 1072 ACCCTAACACAGGCGATGAGAGAAAGTATTTTCAATTTTATCACCAGTACAGACA 1131
Db 481 ACATTTGAGCCCAACCGAAACAGAACTAGATTTTCACTCAACTAAATTAACCCGTTTCGTACA 540
QY 1132 AA 1133
Db 541 AA 542

```

```

RESULT 9
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

```

Query Match 5.2%; Score 59.2; DB 4; Length 640681; Best Local Similarity 44.2%; Pred. No. 0.00015; Matches 455; Conservative 0; Mismatches 548; Indels 27; Gaps 4;									
Qy	4	ATTCAATTTTCAATTAATCGGACATATTTATTCATGCTTTAAATACAACTAAAGTGCT	63						
Db	12549	ATGAAATTTATTAATCAATAATAATTTAAATTTAAATTTTACAAAAATTAAGTCGGTTG	12490						
Qy	64	ATTAGCACTAAATGCCATTCCTATTCCTTCATCAATAAAATTTGAAGTCACITCTACA	123						
Db	12489	CTTGTAAAAATACTTCTTCCTATTTTATGATATGTTATTAATAATATAAAACGGA	12430						
Qy	124	GGAGTAATTTAAACAGGCTCTAAACGGTCAAAATATCAATTTGAAACACATTTCTCTGTAAGT	183						
Db	12429	ATGTTATCATTTAAACAGGACGAAATTTAGAAATAGAAATTTAGTGTATGATTTCAACTATCA	12370						
Qy	184	AATGAAATGCTGGTTGCTTAATACCTCTCCAGAGCTATTTTATTAAGAAGCTAGTTTT	243						
Db	12369	ACAGAACATATCAGGAAGCTGCAACGATTTTCAGCCGCAAACTTTTGA-----T	12319						
Qy	244	TTTATTAATATTTATTTCAAGTTTGCAGATATTTAGTATAAATGTTTAAAGAAATGAACAA	303						
Db	12318	ATTTCAGAAATTCATTAATATTCATCAATATAGAGATGCAATTAACATTAATAATG	12259						
Qy	304	CACCAAGTTGTTTAAACAGTGGTAAATCAGAGATTTACCTTTAAAGGAAAGATGTTGAC	363						
Db	12258	CATATTTCTGGAATAGTCTGTATATATTAACCACTTACCTTATGATGTTTCCCA	12199						
Qy	364	CAGTATCTGCTACAGAGATGATCAACAGAAATCTTTGATTTTAAACAAATTA	423						
Db	12198	GTTTCATCATGCTTTTCATCATATTTTCAGAA-----TTTTTTATACCTTCAGATAT	12148						
Qy	424	TTGAAGTCTATTTATTTGCTGAAACAGCTTTTGCAGCAGTTTACAGAAAGTCTGCTATT	483						
Db	12147	TTAAAAAAGATAGAAAAAATTCATTTTCTATGGCTAAACAGATGTAGTACTAC	12088						
Qy	484	TTAACAGAGTTCAATTTGTTATTAAGTAAATCATAAAGATTTTAAAGCAGTAGCGCTGAC	543						
Db	12087	CTTAATGCTATTTTATTAGAAAAAACAGATCGATCA---CTTATGCAAGTACCCACAGAT	12031						
Qy	544	TCTCATGCTATGAGCAAGTTTAACTTTGGACATATCTCAGAGATTTGATGTA	603						
Db	12030	GGATATGCTCAGGAATATCAAAATTTTTTTTAAAGAAAAATATAATTTCTTTTCAATA	11971						
Qy	604	GTTCTTCCAAGTAAATCTTTGAGAGAAATTTTCAGCAGTATTTACAGATGATTTGAGACC	663						
Db	11970	GTTATTCAGAAAGGCGTTATTGAACATATAGATTTATAATTTCCAAACACCA	11911						
Qy	664	GTTGAGGTATTTTCTCAACCAAGCCAAATCTGTTTCAGAGTGAACACATTTCTTTTAT	723						
Db	11910	ATAAAGTTTATGTTGAAAAAATAATATTAGATFACATATAGAGATCTTATATTACA	11851						
Qy	724	ACAGCCTCTTAGAGGAATTTATCCCATACAGACCGTTTATTAATGACAGATTTGAG	783						
Db	11850	ACAAATTAATTGAAGGCAATATCTGATTATAAGAGTGTGTTGCTTGAATAAATAAAC	11791						
Qy	784	ACGAGGTGTTTTCATACCAATCCCTCTCGCCACGCTATGGAAGTGCCTCTTGATT	843						
Db	11790	AATTTTATTTAAATTTCAAACTATTAAAGCAATCATTTGTACGAGCTGCTATTTA	11731						
Qy	844	TCTAATGCTACTCAAAATGGTACTGTTAAGCTTGAGATTTACTCAAAATCATATTTACGT	903						
Db	11730	TCCCATGAAAAATTTTGTGGAGTGAATAATCATATCAGAAATGGTCAATTTAAAGTATTA	11671						
Qy	904	CATGTTAACTCACCTGAGTGGTGAAGTTAAACAGGAGTTTAGATTTGTTAGTCAGTCT	963						
Db	11670	TCTGATATCAAGAAGAG-----AATAGCAGAAGATAGATTTAATATAATTAATTAAT	11617						
Qy	964	GGTAGTGAATTAACATATCAGCTTCAATCCAACTTACTTTATGAGCTTTTAAAGCTATT	1023						
Db	11616	GGAAATACAGTAAAAATATCAATTTAATGTGTACTACATAATAGAAATATTTAAATTCATT	11557						
Qy	1024	AAAAGTGAAG	1033						
RESULT 10									
US-09-806-708B-22									
; Sequence 22, Application US/09806708B									
; Patent No. 6784342									
; GENERAL INFORMATION:									
; APPLICANT: The University of British Columbia									
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants									
; FILE REFERENCE: 4810-58741									
; CURRENT APPLICATION NUMBER: US/09/806.708B									
; PRIOR FILING DATE: 2001-04-03									
; PRIOR APPLICATION NUMBER: US 60/147,133									
; PRIOR FILING DATE: 1999-08-04									
; NUMBER OF SEQ ID NOS: 23									
; SOFTWARE: Patent in version 3.0									
; SEQ ID NO 22									
; LENGTH: 1141									
; TYPE: DNA									
; ORGANISM: Artificial sequence									
; FEATURE:									
; NAME/KEY: promoter									
; LOCATION: (1)..(1141)									
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters									
US-09-806-708B-22									
Query Match 5.0%; Score 57.2; DB 4; Length 1141; Best Local Similarity 10.2%; Pred. No. 4.7e-05; Matches 110; Conservative 390; Mismatches 571; Indels 5; Gaps 1;									
Qy	58	CGTGCTATTAGCACTAAAAATGCCATTCCTATTCCTTCATCAATAAAAAATTTGAAGTCACT	117						
Db	53	MSKSRKWTWARMYCKYRRWYNKSRWKGWYKKWYBCANNYSRYHARRWKDKMTAYBM	112						
Qy	118	TCTACAGAGTAACTTTAAACAGGCTCTAACCGGTCAAAATATCAATATGAAAAACACTATTCCT	177						
Db	113	TMINKGKTGWRHYWRWRAABDVTVDHHYVTAMNNAWTTTCMMDDKDKRTRWKKNNNA	172						
Qy	178	GTAAGTAATGAAATGCTGGTTTCTTAATACCTCTCCAGGAGCTATTTATTATAGAAGCT	237						
Db	173	TGWDHDDTKYHWNNGCBVTVMVRYKTDNRWSB-----KMNNGMBWKKWSYDVTYTW	227						
Qy	238	AGTTTATTTTATTAATATTTTCAAGTTTGCAGATATAGTATAAATGTTTAAAGAAAT	297						
Db	228	WVDDMKRKYRVRVIRGRMBYVAVBTAHRRYNNGWTEBAMAYRWTNNNNNNNA	287						
Qy	298	GAAACAAACCAAGTTGTTTAAACAGGCTGTAATATCAGAGATTAACCTTAAAGGAAAGAT	357						
Db	288	MCKRAKYWGNRABVNSTCTTWKSKTTKVRTSCWANNCRAGDANKDKHWKWSAAMGVYW	347						
Qy	358	GTTGACCAAGTATCTCGTCTACAGAGATATCAACAGAAATCCCTTGATTTTAAACAA	417						
Db	348	NNNNNNWYTKARHBAWDMVWSHSAWKHANAHSYRKKWTBYKRTMNNNNNGTTTW	407						
Qy	418	AAATTTATGAAGTCTATTATTGCTGAAACAGCTTTTGCAGCCAGTTTCAACAGAAAGCTG	477						
Db	408	KRMWANYKMDMDWEGIVNNNNNGRTYVGTWTKWMTYKWKANNCKWRADHDKTCTH	467						
Qy	478	CCTATTTTAAACAGAGTTTCATATTGTTATTAAGTAAATCAATAAGATTTTAAAGCAGTAGCG	537						
Db	468	NNTTWKMKTWNNCKWSMTNGSHRBAAAVYTYWMMWRRYAHANNNNNNNNNN	527						
Qy	538	ACTGACTCTCATCTATGAGCAACAGTTTAACTTCTTGACAAATCTTTCAGCAGATTTG	597						
Db	528	YKBYCSKWNWNYAANTYKSSWNTSYRYKWTNSWHRSDTRSMGRANNYARABHYGY	587						
Qy	598	ATGAGTAGTCTTCCAAGTAAATCTTTGAGAGAAATTTTTCAGCAGTATTTTACAGATGATTT	657						
Db	588	KWNRTRWBSHTWBHBRAGAAHYWBMWYBAKCHMKAWYKAKKYAGAGGSSNNNNNN	647						
Qy	658	GAGCCGTTGAGGTATTTTCTCCACCAAGCCAAATCTTGTTCAGAGGTGAACACATTTCT	717						



Db 648 NNNNNNATCADDYYAASRWYAMAKWYKYBAANNAYTHANNWGCWNATDTRRT 707  
Qy 718 TTTTATACAGCCTCTTAGAGGAAATATCCGATACAGACCGTTTATTATGACAGAA 777  
Db 708 MWKNNNNNAGTWKNNNNNNAKNAKNAKNAKNAKNAKNAKNAKNAKNAKNAKNAK 767  
Qy 778 TTTGACAGGAGGTTGTTTCAATACCAATCCCTTCGCCAGCTATGGAAGCGCTTC 837  
Db 768 TTDKENNGAYTKYTTNNNNNTYGVVNTAARDGHANNNNNNNNNNNNNNNNNNNN 827  
Qy 838 TTGATTTCTAATGCTACTCAAAATCGTACTGTAAAGCTTGAGATTACTCAAAATCATATT 897  
Db 828 AYANGTNNNNNNNNNAYAMWNTKWTYTTDDWRBAYTNNNNNNNNNNNNNNNNNN 887  
Qy 898 TCAGCTCATGTTAACTCACCTGAGGTGGTAAAGTAAACGAGATTTAGATATTGTTAGT 957  
Db 888 MSDTCDAMKWDATKMNATTYNRGTAWRTNNNNNNNNNNNNNNNNNNNNNNNNNN 947  
Qy 958 CAGTCTGGTAGGATTTAACTATCAGCTTCAATCCAACTTACCTTATTGAGTCTTTAAA 1017  
Db 948 AHTWVCKATKTTKGMNCTTCRKYKNNCTWYTWMTTTRTTWYAATRWKTNATGSM 1007  
Qy 1018 GCTATTAAAGTGAACAGTAAATTCATTTCTTATCACCAGTTTCGACCATTCACCCTA 1077  
Db 1008 RCNATGKNNNNYTWGKTWTAIRWATRMKMAWKVMAATGSMNTNSYARWAYKTRAYKG 1067  
Qy 1078 ACACGAGCGATGAGGAGGAAAGTTTATCCAAATTAATACACCAAGTACGACAA 1133  
Db 1068 YNACAWRWGKATCYMTDANAWTACATSNWATHKYNNWHKCKNNNNNNNNNNNN 1123

## RESULT 11

US-09-806-708B-22/c  
; Sequence 22, Application US/09806708B  
; Patent No. 6784342  
; GENERAL INFORMATION:  
; APPLICANT: The University of British Columbia  
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants  
; FILE REFERENCE: 4810-58741  
; CURRENT APPLICATION NUMBER: US/09/806,708B  
; CURRENT FILING DATE: 2001-04-03  
; PRIOR APPLICATION NUMBER: US 60/147,133  
; PRIOR FILING DATE: 1999-08-04  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 22  
; LENGTH: 1141  
; TYPE: DNA  
; ORGANISM: Artificial sequence  
; FEATURE:  
; NAME/KEY: promoter  
; LOCATION: (1)..(1141)  
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters  
US-09-806-708B-22

Query Match 4.8%; Score 54.8; DB 4; Length 1141;  
Best Local Similarity 12.4%; Pred. No. 0.00019;  
Matches 123; Conservative 335; Mismatches 530; Indels 5; Gaps 2;  
Qy 130 ACTTTAAGCGGTCTAAGCGTCAATATCAATGAAACACATATTCCTGTAAGTAATGAA 189  
Db 1120 ANNNNNNNNGKDWNRMDATKWSATGAWTTHAKRGATMCWYWTGTNRRCMRTYA 1061  
Qy 190 AATGCTGTTGCTAATTAACCTCTCCAGGAGCTATTTATTAGAGCTAGTTTATT 249  
Db 1060 MRTWYTRSNANWSCATKBNWMTWKYATKYRTAWYAMWCMRNMMWCATNGYAKSCATN 1001  
Qy 250 AATATATTATCAAGTTGCGCATATTAGTATAAATGTTAAAGAAATGAACACACCAA 309  
Db 1000 NAMWATTWTAAYAAKAWARWAGNNRMWYGAAGKNGKGCMAAMATMGWADTAGKCN 941  
Qy 310 GTTGTGTTAACCAGGTGGTAAATCAGAGATTACCTTTAAAGGAAAAGATGTTGACCAT 369

Db 940 NNNNNNTDVRNMAKAKNNNNNNNNNAIWTACYNRAATNNKQATHMKWTHGAHSKRTRHT 881  
Qy 370 CCTCGTCTCAAGAAGTATCAACAGAAATCCTTTGATTATTTAAAAACAAAATATTGAG 429  
Db 880 RTCRTKYNNNNNNNARTVYVYHHAARWMAWRTTNNNNNNNNNNNNNNNNNNNNNN 821  
Qy 430 TCTATTATGCTGAAACAGCTTTTGCACGAGTTTACAAGAAAGTCGTCCTATTATAACA 489  
Db 820 SWCNNNNNNNNNNNNNNTWCHYTTANBBCYRANNNNAAARMARTCNMYMEAAVTTTHTD 761  
Qy 490 GGAGTTCATATTGATTAAGTAATCATATAAGATTTTAAAGCAGTACGACTCTCAT 549  
Db 760 WCYTKMNTWYDMMTMTBTTTTRNMTTSTNNNNNNNNNNNNNNNNNNNNNNNNNNNN 701  
Qy 550 CGTATGACCCAAACGTTT---AATCACCTTTGCAACAATCTCTCAGCAGATTTGATCGTAGT 605  
Db 700 NNWGCWNNNTDARRNTTVMRBRWMTNTKRWSTTRHHYTGATNNNNNNNNNNNNNN 641  
Qy 606 TCTTCCAAAGTAAATCTTTGAGAGAAATTTTACAGAGTATTTACAGATGATATTGAGACCGT 665  
Db 640 NNSCTCTRMMTMRWTKGDMTVRKVKVRDITCTYVDVWADSWVWYVANNMRCRDVY 581  
Qy 666 TGAGGTATTTTTCACCAACGCAAACTCTGTTGAGAAGTGAACACATTTCTTTTATAC 725  
Db 580 TRNNTYKSYAHSYWYWSNNNAWYRYSARNWSSMARWTTTRNNNNNNNNNNNNNNNN 521  
Qy 726 ACAGCTCTTAGAGGAAATATCCCGATACAGACCGTTTATTAATGACAGAAATTTGAGAC 785  
Db 520 RHNNNNNTDTRYVWVWKKWARBTITVYDSMCNAKSMWRGNWRAKMKWMAANNNDAGAM 461  
Qy 786 GGAGGTGTTTCAATACCCCAATCCCTTCGCCACGCTATGGAAGCGTCTCTTGAATTC 845  
Db 460 WTYWNGNNTMMRRAMKMMNAWCKRAYCCNNNNNNRACVWHKHKMWRWTKYMWKAA 401  
Qy 846 TAATGCTACTCAAAATGCTACTGTTAAGCTTGAGATTACTCAAAATCATATTTCAGCTCA 905  
Db 400 NNBAKMYRMVAMMYSRDITNTDMMMTSDWBWHYTVDTYTMRAWNNNNNNNNNNNN 341  
Qy 906 TGTTAACTCACCTGAGGTGTTAAGTAAACGAGGATTTAGATATTG-TTAGTCACTG 964  
Db 340 SWMMMDHNTCHTYGNTWGSAYBMAAMSMAAGASBNVYNNWCRWYMGKTWNNNN 281  
Qy 965 GTAGTGATTTAACTATCAGCTTCAATCCAACTTACCTTATTGAGTCTTTAAAGCTATTA 1024  
Db 280 NKAWYRTKTVAWNNRYVDTAVWTBKRYKYCYAYBYWYBYWYMGKHHWBRABHR 221  
Qy 1025 AAAGTGAACAGTAAATAATTCATTTCTTATCACCAGTTCGACCATTCACCCTAACAC 1084  
Db 220 SWNNWVVKCRNKYVSWHYHAMRYBKWABAGCNNNNNNNNNNNNNNNNNNNNNNNN 161  
Qy 1085 CGGATGAGGAGGAAAGTTTATCCAAATTAATTA 1117  
Db 160 MHHKKGKAAWNTKNTABRDDHBAHVKYTYW 128

## RESULT 12

US-09-134-000C-1635  
; Sequence 1635, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; CURRENT FILING DATE: 1998-08-13  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1635  
; LENGTH: 426



```
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-1635

Query Match      4.4%; Score 49.6; DB 4; Length 426;
Best Local Similarity 62.5%; Pred. No. 0.0027;
Matches 95; Conservative 0; Mismatches 54; Indels 3; Gaps 1;

Qy 421 TTATTGAAGTCTATTATTGCTGAACAGCTTTTGCAGCCAGTTTACAAAGAAAGTCTCT 480
Db 259 TTGGGAGAAAATATTTTGGAAACCGGGTGGTTCATGCCCGAAAGTCGCCCC 318
Qy 481 ATTTAACAGAGGTTCATATTGTATTAAAGTAATCATAAAGATTTTAAAGCAGTAGCCGACT 540
Db 319 ATTTAACTGGGTTCACTTTTATTGAGAAATCAAAAATTACTT---GCCGTTGCGACA 375
Qy 541 GACTCTCATCTATGAGCCCAAGCTTTAATCAC 572
Db 376 GATTACATCGTTTAAAGTCAACGTTGTATCCC 407

RESULT 13
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, NASHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match      4.3%; Score 49; DB 4; Length 640681;
Best Local Similarity 53.4%; Pred. No. 0.054;
Matches 103; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

Qy 246 TATTAATATTATTTCAAGTTTGCAGATATTAGTATAAATGTTAAAGAAATTTGAACACAA 305
Db 566908 TAATATTATTTCTTTGAAATAAATAAATAATCATTTTAACTCCAAATAACATGAAATGTT 566967
Qy 306 CCAAGTTGTTTAAACAGTGGTAATCAGAGATTACCTTAAAGGAAAGATGTTGACCA 365
Db 566968 AAATATTATTATTAATCAATATATTTCTATATAGTATATTTTAAATAAATTTTAAATA 567027
Qy 366 GTATCCTCGTCTACAAGAGTATCAACAGAAATCCTTTGATTTTAAATAAATAATTTATT 425
Db 567028 TTTTAAATAACAAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 567087
Qy 426 GAAGTCTATTATT 438
Db 567088 TCATTTTATTATT 567100

RESULT 14
US-09-601-198-56/c
; Sequence 56, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Casseil, Gail H.
; APPLICANT: Chen, Ellison Y.
; APPLICANT: Glass, Jennifer S.
```

```
; APPLICANT: Glass, John I.
; APPLICANT: Heiner, Cheryl R.
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
; TITLE OF INVENTION: UREALYTICUM
; FILE REFERENCE: UAB-13452/22
; CURRENT APPLICATION NUMBER: US/09/601,198
; PRIOR FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 56
; LENGTH: 14066
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-09-601-198-56

Query Match      4.2%; Score 47.4; DB 4; Length 14066;
Best Local Similarity 47.5%; Pred. No. 0.034;
Matches 141; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

Qy 3 GATTCAATTTTCAATTAATCGCACATTTATTTCATGCTTTTAATACAACTAAACGTCG 62
Db 2575 GATTAAATGATAAAGAAAATATAAGTTTAAATTAATGAAACTGGAATCCGGTTTAAAGT 2516
Qy 63 TATTAGCACTAAAAATGCCATTCCTTCTTCTTTCATCAATAAAAAATTTGAAGTCATCTTAC 122
Db 2515 TATTCAAACTCAAAATGATCTATTATGATCTCAACAACATTAATGTAAGTTATC 2456
Qy 123 AGGAGTAACCTTTAACAGGGTCTAACGGTCAAAATATCAATTTGAAAAACACTATTCCTGTAAG 182
Db 2455 AGGAGTTAACTCTAAATATAATGAGCGTCAGATTAAAGTAGCTTTTATAAAGATAATAATA 2396
Qy 183 TAATGAAAATGCTGGTTGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 242
Db 2395 TGTATCTATGAAAGTTCTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2336
Qy 243 TTTTATTAATTAATTTCAAGTTTGCAGATATTAGTATTAATTAATTAATTAATTAATTAAT 299
Db 2335 GTTATCAATTTAAATTCAAATCGAGATATAGTTTGAATAAATTTGAAATAAATCA 2279

RESULT 15
US-09-543-681A-1232
; Sequence 1232, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETTON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 1232
; LENGTH: 471
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-1232

Query Match      4.1%; Score 46.4; DB 4; Length 471;
Best Local Similarity 57.6%; Pred. No. 0.018;
Matches 83; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 159 AATTGAAACAACTATTCCTGTAAGTAATGAAATGCTGGTTTGCCTTAATTAATTAATTAAT 218
Db 3 AATGGGAAGTAATGAGTATAAATTTAGACGAAAATATTATACCTGTAAGTCTATTGG 62
Qy 219 AGCTATTTTATAGAGCTAGTCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 278
```

Db 63 TAGTATTTCAAGTTGGCAGATATTAGTATAAATGTTAAAGAAATTTGAACCAACCAAGT 122  
Qy 279 TATAAATGTTAAAGAAATTTGAACA 302  
Db 123 TATAAATGTTATAGAATTTGAAA 146

RESULT 16  
US-08-956-171E-892  
; Sequence 892, Application US/08956171E  
; Patent No. 6593114  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; Gil H. Choi  
; Patrick S. Dillon  
; Craig A. Rosen  
; Steven C. Barash  
; Michael R. Fannon  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5256  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/956,171E  
; FILING DATE: 20-Oct-1997  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/009,861  
; FILING DATE: January 5, 1996  
; APPLICATION NUMBER: 08/781,986  
; FILING DATE: January 3, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mark J. Hymen  
; REGISTRATION NUMBER: 46,789  
; REFERENCE/DOCKET NUMBER: PB248P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (240) 314-1224  
; TELEFAX: (301) 309-8439  
; INFORMATION FOR SEQ ID NO: 892:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 751 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 892:

US-08-956-171E-892

Query Match 4.0%; Score 45.8; DB 4; Length 751;  
Best Local Similarity 48.6%; Pred. No. 0.03;  
Matches 125; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

Qy 72 TAAAAATGCCATTCCTATTCTTTTCATCAATAAATAAATGGAAGTCATCTTCAGGAGTAAC 131  
Db 290 TACAAGGCAAGCATATCGTCTCTCAATTTGGATTTCGATTGAAACATAAAAAAGATGC 349

Qy 132 TTTAACAGGGTCTAACGGTCAATATCAATTTGAAACACTATTCTGTGAAGTAATGAAA 191  
Db 350 TTTAGCATTTAGAAAAGCGAAAAATAAAGTTGATAAATCTATTGAAACAAGAGTGAAGC 409

Qy 192 TGCCTGGTTTGCTAAATTAACCTCCAGGAGCTATTTTATTAGAAAGCTAGTTTTTTTATTAA 251  
Db 410 GATAGCTCAATATCAAGTTTAAACCGNAATATTATTATTGTAACATCATTTTTTAGGTAT 469

Qy 252 TATTATTTCAAGTTTGGCAGATATTAGTATAAATGTTAAAGAAATTTGAACCAACCAAGT 311  
Db 470 TACATTTCTTGATTGCTGTATGTTGCATTATATATACATAAAGCAATAGATGAAACCGAAGA 529

Qy 312 TGTGTTTAAACCAAGTGGTA 328  
Db 530 TGAGTTAGAGAAATTATA 546

Qy 252 TATTATTTCAAGTTTGGCAGATATTAGTATAAATGTTAAAGAAATTTGAACCAACCAAGT 311  
Db 470 TACATTTCTTGATTGCTGTATGTTGCATTATATATACATAAAGCAATAGATGAAACCGAAGA 529

Qy 312 TGTGTTTAAACCAAGTGGTA 328  
Db 530 TGAGTTAGAGAAATTATA 546

RESULT 17  
US-08-781-986A-892  
; Sequence 892, Application US/08781986A  
; Patent No. 6737248  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,986A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Benson, Bob  
; REGISTRATION NUMBER: 30,446  
; REFERENCE/DOCKET NUMBER: PB248PP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 892:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 751 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; US-08-781-986A-892

Query Match 4.0%; Score 45.8; DB 4; Length 751;  
Best Local Similarity 48.6%; Pred. No. 0.03;  
Matches 125; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

Qy 72 TAAAAATGCCATTCCTATTCTTTTCATCAATAAATAAATGGAAGTCATCTTCAGGAGTAAC 131  
Db 290 TACAAGGCAAGCATATCGTCTCTCAATTTGGATTTCGATTGAAACATAAAAAAGATGC 349

Qy 132 TTTAACAGGGTCTAACGGTCAATATCAATTTGAAACACTATTCTGTGAAGTAATGAAA 191  
Db 350 TTTAGCATTTAGAAAAGCGAAAAATAAAGTTGATAAATCTATTGAAACAAGAGTGAAGC 409

Qy 192 TGCCTGGTTTGCTAAATTAACCTCCAGGAGCTATTTTATTAGAAAGCTAGTTTTTTTATTAA 251  
Db 410 GATAGCTCAATATCAAGTTTAAACCGNAATATTATTATTGTAACATCATTTTTTAGGTAT 469

Qy 252 TATTATTTCAAGTTTGGCAGATATTAGTATAAATGTTAAAGAAATTTGAACCAACCAAGT 311  
Db 470 TACATTTCTTGATTGCTGTATGTTGCATTATATATACATAAAGCAATAGATGAAACCGAAGA 529

Qy 312 TGTGTTTAAACCAAGTGGTA 328  
Db 530 TGAGTTAGAGAAATTATA 546



```
; Sequence 62, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Cassell, Gail H.
; APPLICANT: Chen, Ellison Y.
; APPLICANT: Glass, Jennifer S.
; APPLICANT: Glass, John I.
; APPLICANT: Heiner, Cheryl R.
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
; FILE REFERENCE: UAB-13452/22
; CURRENT APPLICATION NUMBER: US/09/601,198
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 62
; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-09-601-198-62

Query Match 4.0%; Score 45; DB 4; Length 1134;
Best Local Similarity 44.3%; Pred. No. 0.055;
Matches 183; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

QY 60 TCGTATTAGCACTAAATAATGCCATCTCTTATCTTCATCAATAAAATTTGAAGTCACCTTC 119
DB 639 TGAATTTAAATAGTAATTTTGTGTTAGATGTTTAAATATCAACCGTAAATATCTTTT 698

QY 120 TACAGAGTAACCTTTAACAGGCTCTAACGGTCAATATCAATTTGAAACACACTATTCCTGT 179
DB 699 TAAAGAAATTAATTTTGAAGCAAAATGATTTAAATTTAACTAATGCGCAACTAAAAAT 758

QY 180 AGTAATGAATGCTGTTGCTTAATACCTCTCCAGAGCTATTTTATAGAGCTAG 239
DB 759 AACTACTAATAATAGCTAAGTTTCTACACAGCTCCAAACCTGTTGAAATAAAAAGTGT 818

QY 240 TTTTATTTTAAATATTTTCAAGTTTGCAGATATTAGTATAAATGTTAAAGAAATGA 299
DB 819 TTTAATGATACAAATTTGATAATCATCCAGTAGTTTAAATTAATTTGAAGTTTAATGA 878

QY 300 ACAACACCAAGTTGTTTAAACAGTGTGTAATACAGAGATTACCTTTAAAGGAAAGATGT 359
DB 879 TAGTCAGAAATAATTTAAAGAAATGATATTTTAAATAATTAATAAAGAGTAGGACC 938

QY 360 TGACAGATATCTCTCTACAGAGATATCAACAGAAATCCTTTGATTTTAAACAA 419
DB 939 AAATGAAGTCATCTTTTGAAGAAACAGTTTCGATTAGCTAATAATTTTGAAGTATTGTA 998

QY 420 ATTATTGAAGTCTATTATTGCTGAAACAGCTTTTGCAGCCAGCTTTTACAGAA 472
DB 999 AATTGAAATACAAAAAGAAATGAAATAATGAAATCTTAAGTTTAGAATCAA 1051

RESULT 21
US-09-710-279-1771
; Sequence 1771, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1771
; LENGTH: 423
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-1771

Query Match 4.0%; Score 44.8; DB 4; Length 423;
Best Local Similarity 46.5%; Pred. No. 0.056;
Matches 145; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 4 ATTCAATTTTCAATTAATCGCACATATTATTATTCATGCTTTTAAATACAACTAAAGCTGT 63
DB 151 ATTTACTCTCTTATCAAAACGTATATTTAATGTTGAAGTTGAAATTTTAGTTAGAAAAAG 210

QY 64 ATTAGCACTAAAAATGCCATTCCTTCTTCATCAATAAAAAATTTGAAGTCACTTCTACA 123
DB 211 ATGAAATGAAAAAACAATAATTTATATATGTCGAACAAGATGTTAGCGAAGAAATA 270

QY 124 GGAGTAATCTTTAACAGGGTCTAAACGGTCAAAATATCAATTTGAAAAACACTATTCTCTGAAGT 183
DB 271 CTAAATGATTTAGGAATTTTAAAAAGGGAGTTTTTACTCAGATATTGATCCGATATG 330

QY 184 AATGAAATGCTGGTTTGTCTAATTAACCTCTCCAGAGCTATTTTATTTAGAACCTAGTTT 243
DB 331 ATTTAAGATGATGAAATGAAAGAGTATTTTAAAGAGGGCTTTCTTAGCAGGTGTTCT 390

QY 244 TTTATTAATATATTTCAGTTTGCAGATATTAGTATATAATTTTAAAGAAATTTGAACAA 303
DB 423 TTTTATTAATATATTTCAGTTTGCAGATATTAGTATATAATTTTAAAGAAATTTGAACAA 303
```

Db 391 GTAAATAAATCTCGAACAATCTTCATATCATCTTGAAATTTTTCACAATATGAAGATCAT 450  
QY 304 CACCAAGTTGTT 315  
Db 451 TCCGAAGGTCCT 462

RESULT 23  
US-09-134-001C-2705  
; Sequence 2705, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 2705  
; LENGTH: 984  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2705

Query Match 4.0%; Score 44.8; DB 3; Length 984;  
Best Local Similarity 46.5%; Pred. No. 0.058;  
Matches 145; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 4 ATTCAATTTTCAATTAATCGCACATTATTATTTCATGCTTTTAAATACAACTAAACGTCCT 63  
Db 208 ATTACTCTCTTACCAACGATATTTTAATGTGAAGTTGAAATTTTAGTAGAAAAAG 267

QY 64 ATTAGCATAAAATGCCATTCCTATTCTTCATCAATAAAAAATGAAAGTCACCTCTACA 123  
Db 268 ATGAATTTGAAAAAACAATATTATATATGTCGAACAAGATGTTAGCGAAGAATA 327

QY 124 GGAGTAACCTTTAACAGGCTCTAACGGTCAATATCAATGAAACACTATTCCTGTAGT 183  
Db 328 CTAATGATTTAGGAATTTTAAAAAGGGAGTTTCTACTCAGATATTGATCCGGATATG 387

QY 184 AATGAAATGCTGTTGCTTAATACCTCTCCAGGAGCTATTTTATTAGAGCTAGTTT 243  
Db 388 ATTAAGATGATGAATGAAGAAGATTATTAGAGGGGCTTTCTTAGCAGGTTGTTCT 447

QY 244 TTTTATTAATATTTCAGGTTTCCAGATATTAGTATAAATGTTAAAGAAATGAAACAA 303  
Db 448 GTAAATAATCCTGAAACATCTTCATATCATCTCGAAATTTTTCACAATATGAAGATCAT 507

QY 304 CACCAAGTTGTT 315  
Db 508 TCCGAAGGTCCT 519

RESULT 24  
US-09-710-279-3820  
; Sequence 3820, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 3820

Query Match 4.0%; Score 44.8; DB 4; Length 3801;  
Best Local Similarity 50.5%; Pred. No. 0.095;  
Matches 109; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 242 TTTTATTAATATTATTTCAGGTTTCCAGATATTAGTATAAATGTTAAAGAAATGTAAC 301  
Db 3744 TTTTGAACCTTTATAGATGAATACTATAGTTACTTTAATACAGTTGTTAAATGTTGAAG 3685

QY 302 AACCAACAGTTGTTTAAACAGTGTTAAATCAGAGTATACCTTAAAGGAAAGATGTTG 361  
Db 3684 AATCGAACAACAGATGTTTACATGTTAAACAGGACATAGTTTAAACAATGAAGAGAG 3625

QY 362 ACCAGTATCTCTGCTACAGAAGATATCAACAGAAATCTCTTGTGATTTTAAACAAAT 421

; LENGTH: 3315  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: nucleic acid sequence  
US-09-710-279-3820

Query Match 4.0%; Score 44.8; DB 4; Length 3315;  
Best Local Similarity 46.5%; Pred. No. 0.091;  
Matches 145; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 4 ATTCAATTTTCAATTAATCGCACATTATTATTTCATGCTTTTAAATACAACTAAACGTCCT 63  
Db 2611 ATTACTCTCTTATCAACAGTATATTTAATGTGAAGTTGAAATTTTAGTTAGAAAAAG 2670

QY 64 ATTAGCATAAAATGCCATTCCTATTCTTCATCAATAAAAAATGAAAGTCACCTCTACA 123  
Db 2671 ATGAATTTGAAAAAACAATATTATATATGTCGAACAAGATGTTAGCGAAGAATA 2730

QY 124 GGAGTAACCTTTAACAGGCTCTAACGGTCAATATCAATGAAACACTATTCCTGTAGT 183  
Db 2731 CTAATGATTTAGGAATTTTAAAAAGGGAGTTTCTACTCAGATATTGATCCGGATATG 2790

QY 184 AATGAAATGCTGTTGCTTAATACCTCTCCAGGAGCTATTTTATTAGAGCTAGTTT 243  
Db 2791 ATTAAGATGATGAATGAAGAAGATTATTTAAGAGGGGCTTTCTTAGCAGGTTGTTCT 2850

QY 244 TTTTATTAATATTATTTCAGGTTTCCAGATATTAGTATAAATGTTAAAGAAATGAAACAA 303  
Db 2851 GTAAATAATCTGAAACATCTTCATATCACTTGAAATTTTTCACAATATGAAGATCAT 2910

QY 304 CACCAAGTTGTT 315  
Db 2911 TCCGAAGGTCCT 2922

RESULT 25  
US-09-710-279-4271/c  
; Sequence 4271, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 4271  
; LENGTH: 3801  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: nucleic acid sequence  
US-09-710-279-4271

Query Match 4.0%; Score 44.8; DB 4; Length 3801;  
Best Local Similarity 50.5%; Pred. No. 0.095;  
Matches 109; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 242 TTTTATTAATATTATTTCAGGTTTCCAGATATTAGTATAAATGTTAAAGAAATGTAAC 301  
Db 3744 TTTTGAACCTTTATAGATGAATACTATAGTTACTTTAATACAGTTGTTAAATGTTGAAG 3685

QY 302 AACCAACAGTTGTTTAAACAGTGTTAAATCAGAGTATACCTTAAAGGAAAGATGTTG 361  
Db 3684 AATCGAACAACAGATGTTTACATGTTAAACAGGACATAGTTTAAACAATGAAGAGAG 3625

QY 362 ACCAGTATCTCTGCTACAGAAGATATCAACAGAAATCTCTTGTGATTTTAAACAAAT 421

```

Db 3624 ACAAAATTATCCAGTAACTGAACATAATTTGTAGAACAATAATGTAATATGAAGTAAAAA 3565
Qy 422 TATTGAAGTCTATTATGCTGAAACAGACGCTTTTGCAG 457
Db 3564 TTGCACATGCTCTTCTCGAGAACACAGTGCTTTCAG 3529

RESULT 26
US-09-710-279-4300/c
; Sequence 4300, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4300
; LENGTH: 3926
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-4300

Query Match 4.0%; Score 44.8; DB 4; Length 3926;
Best Local Similarity 50.5%; Pred. No. 0.096;
Matches 109; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

Qy 242 TTTTATTAAATATTTTCAAGTTTCCAGATATTAGTATAATGTTAAAGAAATTGAC 301
Db 985 TTTTGAACCTTTATAGATGAATACTATAGTTACTTTAATTAACAGTTGTTAAATGTTGAG 926
Qy 302 AACACCAAGTGTGTTTAAACAGTGGTAAATCAGAGATTACCTTTAAAGGAAAGATGTTG 361
Db 925 ATCGAAACACAGATGTTTACATGTTAAACAGGACATAGTTTAAACAAATGAAGAGAG 866
Qy 362 ACAGTATCCTCGTCTACAGAAAGTATCAACAGAAATCCTTTGATTTTAAAAACAAAT 421
Db 865 ACAAAATTATCCAGTAACTGAACATATTGTTAGAACATAATGTAATATGAAGTAAAAA 806
Qy 422 TATTGAAGTCTATTATTGCTGTAACAGCTTTTTCAG 457
Db 805 TTGCACATGCTCTTCTCGAGAAACAGTGCTTTCAG 770

RESULT 27
US-09-248-796A-778
; Sequence 778, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 778
; LENGTH: 744
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-778
```

```

Query Match 3.9%; Score 43.8; DB 4; Length 744;
Best Local Similarity 46.5%; Pred. No. 0.094;
Matches 141; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

Qy 359 TTGACCAAGTATCCTCGTCTACAAGNAGTATCAACAGAAATCCCTTTGATTTTAAAAACAA 418
Db 320 TAGATCAAAACCCACCTGAAGAAATTTAGAAATATCTGATAATATTAAAGAAATTTAAAAACAA 379
Qy 419 AATTATTGAAGTCTATTATTGCTGAACACAGCTTTTGCAGCCAGTTTACAGAAAGTCGTC 478
Db 380 AATTATTGGAGGAGTGGAGACGATATGAGATGAAATCATTTTGAAGATAGACTG 439
Qy 479 CTATTTTAAACAGGAGTTCATATTGTTAAGTAATATATAAGAAATTTTAAAGCAGTAGCGA 538
Db 440 ATGCTGCTGAAGAGATGACGATGTAATACTAAACCTCTAGTCTAAACGCTGCAAGGAGAT 499
Qy 539 CTGACTCTCATCGTATGAGCCAAACGTTTAAATCACTTTGGACAAATACTTTCAGCAGATTTGA 598
Db 500 CAACCTGGTTCAGGAAAGACAAAGTCGTCACCTGTTCAAGATGCTTTAGAAATTTGGTG 559
Qy 599 TGGTAGTTCTTCCAAAGTAAATCTTTTGAGAGAAATTTTCAGCAGATTTTACAGATGATATTG 658
Db 560 AACCAGAGATATAATAATAATGATTAAGAAATAGAAAGAAATTAATTGAAAGAAAG 619
Qy 659 AGA 661
Db 620 ATA 622

RESULT 28
US-08-956-171E-82
; Sequence 82, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
```

```

;
; LENGTH: 15598 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 82:
US-08-956-171E-82

Query Match      3.8%; Score 43.4; DB 4; Length 15598;
Best Local Similarity 51.3%; Pred. No. 0.36;
Matches 101; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 287 TTAAGAAATTTGAACACACCAAGTGTGTTTAAACCAAGTGGTAAATCAAGATTACCTTAA 346
Db 5165 TGAAGAAATTTGAACACCAAGTGTGTTTAAACCAAGTGGTAAATCAAGATTACCTTAA 5224

QY 347 AAGGAAAAGATGTTGACCAAGTATCCTCGTCTACAAAGAGTATCAACAGAAATCCTTTGA 406
Db 5225 AACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5284

QY 407 TTTTAAAAACAAAATTTTGAAGTCTATTATTGCTGAAACAGCTTTTGCAGCCAGTTTAC 466
Db 5285 TACTTAAACCGTGCACCAACTTCATAGACTTGGTATTCAAGCAATTTGAACCAACTTTAG 5344

QY 467 AAGAAAGTCGTCCTATT 483
Db 5345 TTGAAGGTCGTCGATT 5361
```

```

RESULT 29
US-08-781-986A-82
; Sequence 82, Application US/08781986A
; Patent No. 6737248
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781.986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15598 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-82

Query Match      3.8%; Score 43.4; DB 4; Length 15598;
Best Local Similarity 51.3%; Pred. No. 0.36;
Matches 101; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
```

```

QY 287 TTAAGAAATTTGAACACCAAGTGTGTTTAAACCAAGTGGTAAATCAAGATTACCTTAA 346
Db 5165 TGAAGAAATTTGAACACCAAGTGTGTTTAAACCAAGTGGTAAATCAAGATTACCTTAA 5224

QY 347 AAGGAAAAGATGTTGACCAAGTATCCTCGTCTACAAAGAGTATCAACAGAAATCCTTTGA 406
Db 5225 AACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5284

QY 407 TTTTAAAAACAAAATTTTGAAGTCTATTATTGCTGAAACAGCTTTTGCAGCCAGTTTAC 466
Db 5285 TACTTAAACCGTGCACCAACTTCATAGACTTGGTATTCAAGCAATTTGAACCAACTTTAG 5344

QY 467 AAGAAAGTCGTCCTATT 483
Db 5345 TTGAAGGTCGTCGATT 5361

RESULT 30
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6503729
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916.421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
```



```
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191389)..(191389)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191395)..(191395)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231380)..(231380)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741584)..(741584)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
```

```
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1664854)..(1664854)
OTHER INFORMATION: n equals a, t, c, or g
```

US-08-916-421B-1

```
Query Match      3.8%; Score 43.4; DB 4; Length 1664976;
Best Local Similarity 47.3%; Pred. No. 1.9;
Matches 165; Conservative 0; Mismatches 181; Indels 3; Gaps 1;

QY 73 AAAAATGCCATTCCTTATTCCTTCATCAATAAAAAATGAAGTCACCTTCACAGAGTAAC 132
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1301470 AAAAGAGCAGTTGAATTAGCTAAGCAGCGCAAAAAGAAAACAGTAAAGTTGAGGATAT 1301529

QY 133 TTAACAGGGTCTAACGGTCAATATCAATTGAAACACATTCCTCTGTAAGTAATGAAAT 192
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1301530 AAAATTGGCTTTGAAGAAAATAAATTTTATTTTAAATTTTATTTTATTTATTAATAT 1301589

QY 193 GCTGGTTTGCTAAATACCTCTCCAGGAGCTATTTTATTAGAGCTAGTTTTTTTAA---TT 249
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1301590 TTTACTTCCTCCAAAACCTTTAGATTCCTTTTGTATTATTATTTATTTATTTATTA 1301649

QY 250 AATATTATTCAAGTTTCCAGATATATAGTATAAATGTTAAAGAAATGGAACACACCAA 309
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1301650 TATATTTTAAACCGTGGAAATATGCAAAAAGAAATAAAAAATATAAAATATCGGTA 1301709

QY 310 GTTGTTTTAAACGAGTGGTAATACAGATTACCTTTAAAGGAAAAGATGTTGACACGATAT 369
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



```
NAME/KEY: misc_feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (741584)..(741584)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1095846)..(1095846)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
```

```
Query Match 3.8%; Score 43.4; DB 4; Length 1664976;
Best Local Similarity 47.3%; Pred. No. 1.9;
Matches 165; Conservative 0; Mismatches 181; Indels 3; Gaps 1;

QY 73 AAAAATGCCATCTCTTCTTTTCATCAATAAAAAATGAAAGTCACCTTCTACAGGAGTAAT 132
Db 1301470 AAAAGAAGCAGTGTGAATTAGTAGTAAAGCAACAAAAGAAAACAGTAAGAAGTTGAGGATAT 1301529

QY 133 TTAACAGGCTCTAACCGTCAATATCAATTTGAAAACACTTCTTCGTAAAGTAATGAAAT 192
```

```
Db 1301530 AAAATTCGCTTTGAAGAAATAAATTTTATTTTAAATTTTAAATTTTATTTATTAATAT 1301589
QY 1301593 GCTGGTTTGTCTAAATACCTCTCCAGGAGCTATTTTATTAGAGCTAGTGTCTTTTAA---TT 249
Db 1301590 TTTACTTCCTTCCAAAACCTTTTAGAATTCCTTTGTTTATTTTAAATTTTATTTATTTA 1301649
QY 250 AATATTATTTCAAGTTTCCAGATATTTAGTATAAATGTTAAAGAAATTTGAACAACACCAA 309
Db 1301650 TATATTTTAAACCGTGGAAATATGCACAAAAGAAATATAAAAATATCGGTA 1301709
QY 310 GTTGTGTTTAAACGAGTGAATCAGAGATTACCTTTAAAGGAAAAGAGATGTTGACCAAGTAT 369
Db 1301710 GTTACTGTAAAGTGATAGATATAATGATTTAAATTAAGGAAAAGAGATGATGATATAA 1301769
QY 370 CTTGCTCTACAAGAGTATCAACAGAAAATCCTTTGATTTTAAACCAA 418
Db 1301770 TCTGAAAATTTATTAAGAAAAGAACTAAACGCTAAAGTATATACAATAA 1301818

RESULT 32
US-10-204-708-13/c
; Sequence 13, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 13
; LENGTH: 6113
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-13

Query Match 3.8%; Score 42.8; DB 4; Length 6113;
Best Local Similarity 45.9%; Pred. No. 0.36;
Matches 146; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 5 TTCAATTTTCAATTAATCGCACATTATTTTATTCATGCTTTTAAATACAATAAAGTAACTGCTA 64
Db 1024 TTCTATTACTATTAAATTAATTTTAAATTTCTATTTTAAATACAATAAATAAATACT 965
QY 65 TTAGCACTAAAAATGCCATTCCTATTCTTTCAATCAATAAATAAATGGAAGTCACCTCTACAG 124
Db 964 TTTAAATTAATATATATAATTTCTATAAATTTAAACACATATAATAATAATAACCAACA 905
QY 125 GAGTAACCTTTAAACAGGCTAAACGGTCAAAATATCAATTTGAAAACACTATTTCTGTGAAGTA 184
Db 904 AAATCAAAATACAACAAATTCGCTACCCCTAAATACTCCCTCATACCTCTTTTATA 845
QY 185 ATGAAAATGCTGGTTTGTCTTAATTTACCTCTCCAGGAGCTATTTTATTAGAAGCTAGTTTTT 244
Db 844 ATTACAATCTCTCTCATCCCTAACTCAACCAACCACTAAATTTATTTCTTACCATTATAC 785
QY 245 TTATTAAATTTATTTTCAAGTTTGGCCAGATATTTAGTATAAATGTTAAAGAAATTTGAACAC 304
```

Db 784 TTTTTCATTTAAACTATCCACATATATCTTATACCATATATAACCTTTTAAACCT 725  
 QY 305 ACCAAGTGTGTTTAAACCA 322  
 Db 724 AACTTCTTCTATCAACA 707

RESULT 33  
 US-09-539-333D-208/c  
 ; Sequence 208, Application US/09539333D  
 ; Patent No. 6476208  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cohen, Daniel  
 ; APPLICANT: Blumenfeld, Marta  
 ; APPLICANT: Chumakov, Ilya  
 ; APPLICANT: Bouqueleret, Lydie  
 ; APPLICANT: Bihain, Bernard  
 ; APPLICANT: Essioux, Laurent  
 ; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS  
 ; FILE REFERENCE: GENSET.047AUS  
 ; CURRENT APPLICATION NUMBER: US/09/539,333D  
 ; CURRENT FILING DATE: 2000-03-30  
 ; PRIOR APPLICATION NUMBER: US 60/126,903  
 ; PRIOR FILING DATE: 1999-03-30  
 ; PRIOR APPLICATION NUMBER: US 60/131,971  
 ; PRIOR FILING DATE: 1999-04-30  
 ; PRIOR APPLICATION NUMBER: US 60/132,065  
 ; PRIOR FILING DATE: 1999-04-30  
 ; PRIOR APPLICATION NUMBER: US 60/143,928  
 ; PRIOR FILING DATE: 1999-07-14  
 ; PRIOR APPLICATION NUMBER: US 60/145,915  
 ; PRIOR FILING DATE: 1999-07-27  
 ; PRIOR APPLICATION NUMBER: US 60/146,453  
 ; PRIOR FILING DATE: 1999-07-29  
 ; PRIOR APPLICATION NUMBER: US 60/146,452  
 ; PRIOR FILING DATE: 1999-07-29  
 ; PRIOR APPLICATION NUMBER: US 60/162,288  
 ; PRIOR FILING DATE: 1999-10-28  
 ; PRIOR APPLICATION NUMBER: US 09/416,384  
 ; PRIOR FILING DATE: 1999-10-12  
 ; NUMBER OF SEQ ID NOS: 231  
 ; SOFTWARE: Patent.pm  
 ; SEQ ID NO 208  
 ; LENGTH: 3001  
 ; TYPE: DNA  
 ; ORGANISM: Homo Sapiens  
 ; FEATURE:  
 ; NAME/KEY: allele  
 ; LOCATION: 1501  
 ; OTHER INFORMATION: 99-26772-268 : polymorphic base C or T  
 ; FEATURE:  
 ; NAME/KEY: misc\_binding  
 ; LOCATION: 1481..1500  
 ; OTHER INFORMATION: 99-26772-268.mis1,  
 ; FEATURE:  
 ; NAME/KEY: misc\_binding  
 ; LOCATION: 1502..1520  
 ; OTHER INFORMATION: 99-26772-268.mis2, complement  
 ; FEATURE:  
 ; NAME/KEY: primer\_bind  
 ; LOCATION: 1235..1254  
 ; OTHER INFORMATION: upstream amplification primer  
 ; FEATURE:  
 ; NAME/KEY: primer\_bind  
 ; LOCATION: 1702..1722  
 ; OTHER INFORMATION: downstream amplification primer, complement  
 ; FEATURE:  
 ; NAME/KEY: misc\_binding  
 ; LOCATION: 1489..1513  
 ; OTHER INFORMATION: 99-26772-268 probe  
 ; US-09-539-333D-208

Query Match 3.7%; Score 42.4; DB 4; Length 3001;  
 Best Local Similarity 55.6%; Pred. No. 0.35;  
 Matches 79; Conservative 1; Mismatches 62; Indels 0; Gaps 0;  
 QY 311 TTGTTTTAACCAAGTGGTAAATCAGAGATTACCTTAAAGGAAAGATGTTGACCAAGTATC 370  
 Db 1585 TGGTTTTCAAAATATGGTAATAGAGAAGTAATCTACACAGGGTAAAAAATATACCTAAAT 1526  
 QY 371 CTCGCTACAAGAAGTATCAACAGAAATCCTTTGATTTTAAACAAAAATTTGAAGT 430  
 Db 1525 GAAATATTTTGAATTTTAACTATRAAATCTTGGATTTTACAAATTAATTTTAG 1466  
 QY 431 CTATTATTGCTCAAAACAGCTTT 452  
 Db 1465 TTGATTTTCTGTAGCATTTCT 1444

RESULT 34  
 US-09-248-796A-6131  
 ; Sequence 6131, Application US/09248796A  
 ; Patent No. 6747137  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Keith Weinstock et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
 ; FILE REFERENCE: 107196.132  
 ; CURRENT APPLICATION NUMBER: US/09/248,796A  
 ; CURRENT FILING DATE: 1999-02-12  
 ; PRIOR APPLICATION NUMBER: US 60/074,725  
 ; PRIOR FILING DATE: 1998-02-13  
 ; PRIOR APPLICATION NUMBER: US 60/096,409  
 ; PRIOR FILING DATE: 1998-08-13  
 ; NUMBER OF SEQ ID NOS: 28208  
 ; SEQ ID NO 6131  
 ; LENGTH: 2919  
 ; TYPE: DNA  
 ; ORGANISM: Candida albicans  
 ; US-09-248-796A-6131

Query Match 3.7%; Score 42.2; DB 4; Length 2919;  
 Best Local Similarity 42.6%; Pred. No. 0.39; Mismatches 298; Indels 0; Gaps 0;  
 Matches 221; Conservative 0;  
 QY 9 ATTTTCAATTAATCGCACATTTATTTATTCATGCTTTAAATACAACTAAACGTCGTATTAG 68  
 Db 2289 ATTATTACAATTTGATAAAATTTTCGATAAAGATTTTACTGAAACAAATTTGAATATT 2348  
 QY 69 CACTAAAAATGCCAATTCCTATTCTTTTCATCAATAAAAATTTGAAGTCATCTTACAGAGT 128  
 Db 2349 TGCTGATATTGATAAAGCAAGTTTAAACCATAATTTGATGATATTTTAAATATTTGAAAAAT 2408  
 QY 129 AACTTTAACAGGGTCTAACGGTCAATATCAATTTGAAAAACACTATTCTCTGAAGTAATGA 188  
 Db 2409 ATATATTGCTCAAACTGTCAATTAATTTTCCATACCACTGGTCCAAAGCCAAAGC 2468  
 QY 189 AAATGCTGGTTGCTAATTAACCTCTCCAGGAGCTATTATTATTAGAAGCTAGTTTTTTAT 248  
 Db 2469 CAAAGGAACTAAATATTAGTTATATCTTCAACCAAGTTTAAAAAGTTGATGTTGTTA 2528  
 QY 249 TAATATTATTCAAGTTTGCAGATATTAGTATAATTTGTTAAAGAAATTTGAACCAACCA 308  
 Db 2529 TGAGATTTTAATTCATTTTGTAAACTTATCAATAAGATCAAAACCATTAATGAAGA 2588  
 QY 309 AGTTGTTTTTAAACCAAGTGGTAAATCAGAGATTACCTTAAAGGAAAGAGATGTTGACCAGTA 368  
 Db 2589 AGTATTGTCACAATTTAATGAATTAATGAATTTTGAATAATTTCTTGATAA 2648  
 QY 369 TCCTCGTCTACAAGAGTATCAACAGAAAATCCCTTTGATTTTAAAAACAAAAATTTTGA 428  
 Db 2649 TACTCGTCAAAATACCTTTTAAAACTGTCTATTAGCTTATGTTAAATTTAAATTTAGATGCAA 2708  
 QY 429 GTCTATTATTGCTGAACACAGCTTTTGAGCCAGCTTTTACAGAAAGTGTCTCTATTAAAC 488

```
Db 2709 TTTTATTCTTGTATTTGAAATAACAAACCTTTTACAAATTAATGATCAAGTTATAA 2768
Qy 489 AGAGTTCATATTTGTTAAGTAATCAATAAGATTTTAA 527
Db 2769 AATTTTACAAATTTTATTAAATGAAATTCATAATAAAAA 2807

RESULT 35
US-09-248-796A-5237
; Sequence 5237, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 5237
; LENGTH: 1782
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-5237

Query Match 3.7%; Score 41.6; DB 4; Length 1782;
Best Local Similarity 46.6%; Pred. No. 0.46; Mismatches 189; Indels 1; Gaps 1;
Matches 166; Conservative 0;

Qy 72 TAAAAATGCGCATCTCTTCTTTCATCAATAAAAAATTGAAGTCACCTTCTACAGGAGTAAC 131
Db 831 TCAAGAAATTTACTCCAATTTCTTGACCTTAATAATGTTAATGTCATACTAATGTTTGG 890
Qy 132 TTTACAGGTCCTAAGGTCAAATCAATGGAACACACTATTCCTGTAAGTAAGTAAGAAA 191
Db 891 TGGTACAAGTGTGTGTGATGAGAACCAATACCACAAATAATGATATTTCTTCAAAA 950
Qy 192 TGCTGTTTGCTAAATACCTCTCCAGGAGCTATTTTATTAGAAGCTAGTTTATTATAA 251
Db 951 TCAGAAACATTAATATCTGGCAATTTATTGATGATATCTGAAATTTTATCAATGA 1010
Qy 252 TATTATTTTCAAGTTTCCAGATATTAGTATAAATGTTAAAGAAATTGA-ACAACACCAAG 310
Db 1011 ATCATTAATTTTTCAAAGTTTGGTAGAATTTTAGAAAAATTTGATACTACATCAAC 1070
Qy 311 TTGTTTTAACAGTGTGTAATCAGAGATTACCTTAAAGGAAAGATGTTGACCAATATC 370
Db 1071 CATTTGTTTCAAGTGGGAGTACTAACAATAATAATCAATTAAGAAATTTTATAAATTT 1130
Qy 371 CTCGCTACAAGAATATCAACAGAAATCCCTTGATTTTAAACAAAAATTTATG 426
Db 1131 GCGTGCCACAGCAATAACAACAACAACAACAACAATACTATCATCTAACAGATTG 1186

RESULT 36
US-09-893-600-1/c
; Sequence 1, Application US/09893600
; Patent No. 6759205
; GENERAL INFORMATION:
; APPLICANT: Hansen, J. No. 6759205man
; TITLE OF INVENTION: Construction of a Strain of Bacillus subtilis 168 that Displays B
; FILE OF INVENTION: Sublancin Lantibiotic on the Surface of the Cell
; FILE REFERENCE: 108172-00057
; CURRENT APPLICATION NUMBER: US/09/893,600
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: 60/215,449
; PRIOR FILING DATE: 2000-06-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 1
; LENGTH: 2517
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: The EcoRI-HindIII insert of the pLPVc integrative plasmid.
US-09-893-600-1

Query Match 3.7%; Score 41.6; DB 4; Length 2517;
Best Local Similarity 43.4%; Pred. No. 0.52;
Matches 191; Conservative 0; Mismatches 249; Indels 0; Gaps 0;

Qy 242 TTTTATTAAATTAATTTCAAGTTTGCAGATATAGTATAAATCTTAAAGAAATTTGAC 301
Db 1787 TTTTGGTTTTCGAGTCTCCGAGTTTAACTTCTTTAAATAGCTTTTCCATTTGTAAC 1728
Qy 302 AACCAAGTTGTTTAAACCAAGTGTAAATCAGAGATTACCTTAAAGGAAAGATGTTG 361
Db 1727 CTCCCATTTGTTTGCATTCGGATTACATTTTAAATATAAGGAAGATAAATCTAAAT 1668
Qy 362 ACCAGTATCTCTGCTACAGAAGTATCAACAGAAAAATCTTTGATTTTAAACAAAAT 421
Db 1667 TCAACCATATTTTATTTTATTTTGAATAATATAAGGAAACTTGTTCATAAATCATAGA 1608
Qy 422 TATTGAAGTCTATTATGCTGAACAGCTTTTGCAGCCAGTTTACAAGAAAGTCGCTCA 481
Db 1607 AAAAAAATTTTATTTACCTTATCAATAAACAAGATTTGTTGTAATACACGAATCCCTTCTTC 1548
Qy 482 TTTTAAACAGGAGTTTCATATTGTTTAACTAATATAAAGATTTTAAACAGTAGGACTG 541
Db 1547 AACTAACGGGCGAGGTGTAGTACATTAGAAACCGACTGTAAAGAGTACAGTCGGCATTA 1488
Qy 542 ACTCTCATCTGATGAGCCAACTTTTAAATCACTTTGGACAATATCTTCAGCAGATTTGATGG 601
Db 1487 TCTCATATTATAAAGCCAGTCATTAGGCTATCTGACAAATCTCTGAATAGAGTTTCATAA 1428
Qy 602 TAGTCTTCCAAAGTAATCTTTGAGAGAAATTTTCAGCAGATTTTACAGATGATATTGAGA 661
Db 1427 ACNAATCTGCAATGATACCATCAACAACAGAAATGATGTACCTGTAAAGATAGCGGTAAT 1368
Qy 662 CCGTTGAGGTATTTTCTCA 681
Db 1367 ATATTGAATTAACCTTTATTA 1348

RESULT 37
US-09-248-796A-11532
; Sequence 11532, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 11532
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-11532

Query Match 3.7%; Score 41.4; DB 4; Length 603;
Best Local Similarity 43.8%; Pred. No. 0.35;
Matches 180; Conservative 0; Mismatches 231; Indels 0; Gaps 0;

Qy 159 AATTGAAACACTATTCCTGTAAGTAATGAAATGCTGTTTGTGCTTAATTCCTCTCCAGG 218
```

Db 114 AAGTCCCATCCAGTCTTGATATAAATGTGATTAATAATTTCTTGATACCATCCAA 173  
Qy 219 AGCTATTTTATAGAACTAGTCTTTTATTAATATATATTTCAAGTTGCGAGATATTAG 278  
Db 174 ATGTGATATTGATCAAAATATGTCATTTGTAATAATGAATGAACTGATTAATTTATGTAGGAA 233  
Qy 279 TATAAATGTTTAAAGAAATGAACAACACCAAGTGTGTTTAAACCAAGTGGTAATCAAGAT 338  
Db 234 AAATGATACAAGGGAATGATATGGGACAGAAATTAAGAGAGAGATTTAAGAGAGAG 293  
Qy 339 TACCTTAAAGGAAAGATGTTGACCAAGTATCCTCGTCTACAAGAAAGTATCAACAGAAA 398  
Db 294 TGACTTAAAGTGAATGATATACAACAAATGAGTTAAGCGAAAGCATTTCAAGTGA 353  
Qy 399 TCCTTTGATTTTAAACAAATTTTGAAGTCTATTTATGCTGTAACAGCTTTTCCAGC 458  
Db 354 TGATTTAAGAGAAATAGCAACAAAGAACTGTTAGTGAATTTTGAAGAAATCAATGCA 413  
Qy 459 CAGTTTACAAGAAAGTGTCTCTATTTTAAACAGGAGTTCATATTTGTTAAGTAATCATAA 518  
Db 414 GAATTTGTTTGGACTTGGTCATGATTCATATGATGATGTTCTACCTGGCATTTAGAAACATCA 473  
Qy 519 AGATTTTAAAGCAGTAGCACTGCTCATCTGATGAGCCAGCTTTTAAT 569  
Db 474 GTATTTCCAGATTTATTTCTAGACTTCAGAGTTTTCGTGGAATAGAAAT 524

## RESULT 38

US-09-248-796A-444  
; Sequence 444, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; PRIOR FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 444  
; LENGTH: 825  
; TYPE: DNA  
; ORGANISM: Candida albicans  
US-09-248-796A-444

Query Match 3.7%; Score 41.4; DB 4; Length 825;  
Best Local Similarity 44.1%; Pred. No. 0.39;  
Matches 174; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

Qy 276 TAGTATAATGTTAAAGAAATGAACAACCAAGTCTGTTTAAACCAAGTGGTAATCAGA 335  
Db 132 TAGTCAACACAGAAAGATATACTGATCCAAACAACTAATAACCATATATAAAAAACA 191  
Qy 336 GATTACCTTAAAGGAAAGATGTTGACCAAGTATCCTGCTACAAGAAAGTATCAACAGA 395  
Db 192 AGGTAATTTGATATCAAGACGACTCTTATAGATAATTTCAACCAATCAGAAACTTA 251  
Qy 396 AAATCCTTTGATTTTAAACAAATATTTCAAGTCTATTTTCTGCTGAACAGCTTTTGC 455  
Db 252 CAATTAATTTATTTAAATTTAATTTAATTTGAGAAATAAGTGAACCAAGATCCTAA 311  
Qy 456 AGCCAGTTTACAAGAAAGTCTCTTATTTTAAACAGGAGTTCATATTTGTTAAGTAATCA 515  
Db 312 TATATTAATGAAGAAAGGAGGAAATGCGAGCATTAATACAAGAGAAATGACGACAAA 371  
Qy 516 TAAAGATTTTAAAGCAGTAGCACTGCTCATCTGATGAGCCAAAGCTTTTAACTTT 575  
Db 372 TAATGATAATAATGAATTTATTAATATAGTTGATTAAGATATTTCAAGATAAATATTGA 431

Qy 576 GGACAATACTTCCAGCAGATTTGATGTAGTCTTCCAAAGTAAATCTTTGAGAGAAATTTTC 635  
Db 432 TAGTCTTTGAATTTTCAATAATTTAGTTTAAATAATGATTTAATTTGATTAATAACAGAACTTTT 491  
Qy 636 AGCAGTATTTACAGATGATATTTGAGACCGTTTGAGG 670  
Db 492 AGGTATAAGTATGAAGAAATTTGCTTAAATTTAAAG 526

## RESULT 39

US-09-806-708B-23/c  
; Sequence 23, Application US/09806708B  
; Patent No. 6784342  
; GENERAL INFORMATION:  
; APPLICANT: The University of British Columbia  
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants  
; FILE REFERENCE: 4810-58741  
; CURRENT APPLICATION NUMBER: US/09/806,708B  
; CURRENT FILING DATE: 2001-04-03  
; PRIOR APPLICATION NUMBER: US 60/147,133  
; PRIOR FILING DATE: 1999-08-04  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 23  
; LENGTH: 1055  
; TYPE: DNA  
; ORGANISM: Artificial sequence  
; FEATURE:  
; NAME/KEY: promoter  
; LOCATION: (1)..(1055)  
; OTHER INFORMATION: consensus sequence of A.t. and L.a. PAB1 promoters  
US-09-806-708B-23

Query Match 3.7%; Score 41.4; DB 4; Length 1055;  
Best Local Similarity 20.4%; Pred. No. 0.43;  
Matches 142; Conservative 166; Mismatches 382; Indels 5; Gaps 2;

Qy 223 ATTTTATTAGAGCTAGTCTTTTATTATTAATTAATTTTCAAGTTTGCAGATATTAGTATA 282  
Db 759 AYTWTGTCNN 700  
Qy 283 AATGTTAAAGAAATGAACAACCAAGTGTGTT-TTAACCAAGTGTGTTAAATCAAGATTAC 341  
Db 699 TRCTCTAANTATVTRWATTAATTTTTRWATTTTSTKNTNNNNNNNNNNNNNNNNNNNNNNNN 640  
Qy 342 CTTAAAGGAAAGATGTTGACCAAGTATCCTGCTACAAGAGTATCAACAGAAATCC 401  
Db 639 ATWGCATNNNTWARATNNNTTRAGGAATNTKTCTTSTTAGWYTGATTCGNNNNNNNNNN 580  
Qy 402 TTTGATTTTAAAAACAAAATTTTGAAGTCTATTATTGCTGAAACAGCTTTTGCACCCAG 461  
Db 579 MTASCTCTGACTAATTAATGCGATSAKKTATTTCTYRRRWANGWRTWTANMRCWRT 520  
Qy 462 TTTCAAGAAAGCTGCTCTATTTTTAAACAGAGTTTCATATTGTATTAAAGTAATCATAAAGA 521  
Db 519 YTRTYCKSTAMS---CTYGNAMWYAAVSAGNTSSMARWTTTANNNTAAGYWRWAAWAGTM 464  
Qy 522 TTTTAAAGCAGTAGCAGCTGACTCTCATCGTATGACCAAGCTTTTATCATCTTTTGACAA 581  
Db 463 WAAMANNNTTRYAYWTRWARYTTRYRSACNKSMMWRGATWRAATTTTAANNKAGAM 404  
Qy 582 TACTTCAGCAGATTTGATGTAGTCTTCTCCAGTAATCTTTTGAGAGAAATTTTTCAGCACT 641  
Db 403 WMTTAAAGNNNTTAAATTKMAAWCARVCCNNNNWAAACMAKWKWAWTWKYAWGAACN 344  
Qy 642 ATTACAGATGATTTGAGACCGTTGAGGTATTTTCTCAACAGCCAAATCTTGTTCAG 701  
Db 343 NNNKTAMYCRRAWYSAWTTTWWAAATTSWKWYTTTTRKTTMAAAAANNNNNNNNNNNNNNNNN 284  
Qy 702 AAGTGAACACATTTCTTTTATACAGCTCTTTAGAGAAATTTATCCGATACAGCCG 761  
Db 283 SAWWAWWMAWTCTCGATWGGATYMAAACCTTAAGAGNSMTYWCWAAATYMGTTTNNNNN 224







GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2005, 10:52:06 ; Search time 437 Seconds  
(without alignments)  
13622.067 Million cell updates/sec

Title: US-10-048-071-27  
Perfect score: 1134  
Sequence: 1 atgattcaatttcaattaa.....ttaccacagtagcaacaac 1134

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues  
Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_23Sep04.\*  
1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002as.\*  
7: Geneseqn2002bs.\*  
8: Geneseqn2003as.\*  
9: Geneseqn2003bs.\*  
10: Geneseqn2003cs.\*  
11: Geneseqn2003ds.\*  
12: Geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1134	100.0	1134	4	Aaf54744 Nucleotid
2	1126	99.3	1134	6	Abn68601 Streptoco
3	1126	99.3	1137	6	Abx83362 Streptoco
5	1124.4	99.2	1137	8	Acas0247 Prokaryot
6	849	74.9	3200	2	Abx83363 DNA encod
7	690.8	60.9	1134	8	Aav58285 S. pyogen
8	641.2	56.5	1134	6	Acas0247 Prokaryot
9	641.2	56.5	55361	6	Abn68600 Streptoco
10	608.2	53.6	1134	10	Continuation (22 o
11	608.2	53.6	21338	2	Abx08064 S. pneumo
12	608.2	53.6	110000	10	Aav52153 Streptoco
13	603.4	53.2	1137	4	Continuation (21 o
14	603.4	53.2	1137	8	Aas55711 Streptoco
15	408.2	36.0	110000	6	Acas0247 Prokaryot
16	341.2	30.1	1131	4	Abas0521 Genomic s
17	338	29.8	1128	8	Aas53173 Enterococ
18	330.6	29.2	8001	2	Acas0247 Prokaryot
19	330.6	29.2	8001	6	Aax13559 Enterococ
20	317	28.0	1155	6	Abas0521 Genomic s
21	315.6	27.8	1128	8	Adc91275 E. faeciu
					Acas33953 Prokaryot

22	247.4	21.8	1131	8	ACA46779	AcA46779 Prokaryot
23	247.4	21.8	1158	6	ABN92866	Abn92866 Staphyloc
24	243.4	21.5	1131	8	ACA47854	AcA47854 Prokaryot
25	242	21.3	110000	10	ADF77343_00	Adf77343 Lactic ac
26	236.2	20.8	1131	8	ACF74365	AcF74365 Staphyloc
27	236.2	20.8	1134	4	AAS54991	Aas54991 Staphyloc
28	236.2	20.8	1134	6	ABN86787	Abn86787 Staphyloc
29	236.2	20.8	1134	10	ADD26271	Add26271 Staphyloc
30	236.2	20.8	1134	10	ADD26273	Add26273 Staphyloc
31	236.2	20.8	2347	2	AAV74464	Aav74464 Staphyloc
32	233	20.5	1134	2	Aaz31005	Aaz31005 Partial d
33	233	20.5	1134	4	AAS54965	Aas54965 Staphyloc
34	233	20.5	1134	4	AAS54735	Aas54735 Nucleotid
35	233	20.5	1134	8	ACA20030	AcA20030 Prokaryot
36	233	20.5	1134	10	AAD62920	Aad62920 Staphyloc
37	230	20.3	1128	6	AAS52032	Aas52032 Staphyloc
38	200.4	17.7	4736	6	ABQ70950	Abq70950 Listeria
39	198.8	17.5	110000	6	ABQ67196_2	Continuation (3 of
40	198.8	17.5	110000	6	ABQ69245_00	Abq69245 Listeria
41	186	16.4	110000	6	ABA03041_00	AbA03041 Listeria
42	185.6	16.4	1146	8	ACA36279	AcA36279 Prokaryot
43	181.2	16.0	1218	9	ADB07181	AdB07181 Alloiococ
44	181.2	16.0	1218	9	ADB07183	AdB07183 Alloiococ
45	181.2	16.0	1218	9	ADB07179	AdB07179 Alloiococ

## ALIGNMENTS

RESULT 1  
Aaf54744  
ID Aaf54744 standard; DNA; 1134 BP.

AC AAF54744;

DT 15-MAY-2001 (first entry)

DE Nucleotide sequence of a partial dnaN (beta subunit) polypeptide.

KW dnaB; Gram positive bacteria; polC; dnaE; hola; holB; dnaX; ssb;  
KW dnaG; dnaB; antibiotic; replication; cell growth; cell death;  
KW bacterial infection; ss.

OS Streptococcus pyogenes.

PN WO200109164-A2.

PD 08-FEB-2001.

PF 28-JUL-2000; 2000WO-US020666.

PR 29-JUL-1999; 99US-0146178P.

PA (UVRQ) UNIV ROCKEFELLER.

XX O'donnell ME, Bruck I, Zhang D, Whipple R;

XX WPI; 2001-147453/15.

XX P-PSDB; AAB31944.

PT Isolated DNA molecule from a Gram positive bacterium encoding DNA  
PT replication proteins used to identify compounds which have antibiotic  
PT activity.

PS Claim 38; Page 61-62; 239pp; English.

XX The present sequence encodes a partial dnaN polypeptide. The  
CC specification describes DNA molecules from Gram positive bacteria, which  
CC comprise a coding region from a polC, dnaE, hola, holB, dnaX, ssb,  
CC dnaG or a dnaB gene. These sequences encode proteins that replicate the  
CC chromosome of Gram positive bacteria. They are used for sequencing and  
CC amplification of DNA and in drug discovery to identify compounds which  
CC have antibiotic activity through interference with replication. They are

CC used in methods for identifying compounds that are active at the level of  
CC DNA replication and result in arrest of cell growth or cell death of  
CC bacteria to treat bacterial infections in animals

XX SQ Sequence 1134 BP; 383 A; 201 C; 178 G; 372 T; 0 U; 0 Other;

Query Match 100.0%; Score 1134; DB 4; Length 1134;  
Best Local Similarity 100.0%; Pred. No. 2.1e-232;  
Matches 1134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGATTCATTTTCAATTAATGCGACATTAATTTATTCATGCTTTAAATACAACTAAACGT 60
DB 1 ATGATTCATTTTCAATTAATGCGACATTAATTTATTCATGCTTTAAATACAACTAAACGT 60

QY 61 GCTATTAGCACTAAAATGCCATTCCTATCTTCATCAATAAAAATTTGAAGTCACTTCT 120
DB 61 GCTATTAGCACTAAAATGCCATTCCTATCTTCATCAATAAAAATTTGAAGTCACTTCT 120

QY 121 ACAGGAGTAACCTTTAACAGGGTCTAACGGTCAAAATATCAATTTGAAAAACACTATTCCTGTA 180
DB 121 ACAGGAGTAACCTTTAACAGGGTCTAACGGTCAAAATATCAATTTGAAAAACACTATTCCTGTA 180

QY 181 AGTAATGAAATGCTGGTTGCTTAATTAACCTCCAGGAGCTATTTTATTAGAACCTAGT 240
DB 181 AGTAATGAAATGCTGGTTGCTTAATTAACCTCCAGGAGCTATTTTATTAGAACCTAGT 240

QY 241 TTTTATTATTAATTAATTTCAAGTTTGCAGATATTAGTATAAATGTTAAAGAAATTCGA 300
DB 241 TTTTATTATTAATTAATTTCAAGTTTGCAGATATTAGTATAAATGTTAAAGAAATTCGA 300

QY 301 CAACACCAAGTTGTTTAAACCAAGTGGTAAATCAGAGATTACCTTAAAAAGGAAAGATGTT 360
DB 301 CAACACCAAGTTGTTTAAACCAAGTGGTAAATCAGAGATTACCTTAAAAAGGAAAGATGTT 360

QY 361 GACCAAGTATCCTCGTCTACAGAGATATCAACAGAAATTCCTTTGATTTTAAAAACAAA 420
DB 361 GACCAAGTATCCTCGTCTACAGAGATATCAACAGAAATTCCTTTGATTTTAAAAACAAA 420

QY 421 TTAATTGAAGTCTAATTAATGCTGAAACAGCTTTTGACAGCAGTTTACAAAGAAAGTGGTCT 480
DB 421 TTAATTGAAGTCTAATTAATGCTGAAACAGCTTTTGACAGCAGTTTACAAAGAAAGTGGTCT 480

QY 481 ATTTTAAACAGGAGTTTCAATATTCATTAATTAAGTAAATCAATAAGATTTTAAAGCAGTACGACT 540
DB 481 ATTTTAAACAGGAGTTTCAATATTCATTAATTAAGTAAATCAATAAGATTTTAAAGCAGTACGACT 540

QY 541 GACTCTCATCGTATGAGCCACGTTTAAATCACTTTTGGACATACCTTCAGCAGATTTGATG 600
DB 541 GACTCTCATCGTATGAGCCACGTTTAAATCACTTTTGGACATACCTTCAGCAGATTTGATG 600

QY 601 GTAGTTCTTCCAGTAATATCTTTGAGAGAAATTTTACAGAGATTTTACAGATGATATTGAG 660
DB 601 GTAGTTCTTCCAGTAATATCTTTGAGAGAAATTTTACAGAGATTTTACAGATGATATTGAG 660

QY 661 ACCGTGTAGGATTTTCTTCAACAGCCAAATCTTTGTTTGAAGTGAACACATTTCTTTT 720
DB 661 ACCGTGTAGGATTTTCTTCAACAGCCAAATCTTTGTTTGAAGTGAACACATTTCTTTT 720

QY 721 TATACAGCGCTTTAGAGGAATTTATCCGATACAGACCGTTTATTATGACAGAAATTT 780
DB 721 TATACAGCGCTTTAGAGGAATTTATCCGATACAGACCGTTTATTATGACAGAAATTT 780

QY 781 GAGAGGAGGTTGTTTTCATACCAATCCCTTCGACGCTATGGAAGTGGCTTCTTGTG 840
DB 781 GAGAGGAGGTTGTTTTCATACCAATCCCTTCGACGCTATGGAAGTGGCTTCTTGTG 840

QY 841 ATTCTTAATGCTACTCAAAATGGTACTGTTAAGTTGAGTATCTCAAAATCATATTTCA 900
DB 841 ATTCTTAATGCTACTCAAAATGGTACTGTTAAGTTGAGTATCTCAAAATCATATTTCA 900

QY 901 GCTCATGTTAACTCACTCAGGTTGTTAAGGTAAGGAAATTTAGATATTTGTTAGTCAG 960
DB 901 GCTCATGTTAACTCACTCAGGTTGTTAAGGTAAGGAAATTTAGATATTTGTTAGTCAG 960
```

```
QY 961 TCTGTTAGTGAATTAACCTTACAGCTTCAATCCAACTTACCTTATTGAGTCTTTAAAGCT 1020
DB 961 TCTGTTAGTGAATTAACCTTACAGCTTCAATCCAACTTACCTTATTGAGTCTTTAAAGCT 1020

QY 1021 ATTAAAGTGAAACAGTAATAAATTCATTTCTTATCAGCAGTTCCAGCATTACCTTAACA 1080
DB 1021 ATTAAAGTGAAACAGTAATAAATTCATTTCTTATCAGCAGTTCCAGCATTACCTTAACA 1080

QY 1081 CCAGCGCATGAGGAAGAAAGTTTATCAATTAATTAACACCATGACGAAACAAAC 1134
DB 1081 CCAGCGCATGAGGAAGAAAGTTTATCAATTAATTAACACCATGACGAAACAAAC 1134

RESULT 2
ABN68601
ID ABN68601 standard; DNA; 1134 BP.
XX
AC ABN68601;
XX
DT 01-JUL-2002 (first entry)
XX
DE Streptococcus polynucleotide SEQ ID NO 5115.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS Streptococcus agalactiae.
XX
FN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB004789.
XX
PR 27-OCT-2000; 2000GB-00026333.
PR 24-NOV-2000; 2000GB-00028727.
PR 07-MAR-2001; 2001GB-00005640.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Massignani V, Margarit Y Rosl, Grandi G, Fraser C;
PI Tettelin H;
XX
WPI: 2002-352536/38.
P-PSDB: ABP27970.
XX
New Streptococcus protein for the treatment or prevention of infection or
disease caused by Streptococcus bacteria, such as meningitis, and for
detecting a compound that binds to the protein.
XX
Claim 7; Page 3675; 4525pp; English.
XX
The invention relates to a protein (ABP25413-ABP30895) from group B
streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
(Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
the specification. The proteins have antibacterial and antiinflammatory
activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
antibodies that bind (I) are used in the manufacture of medicaments for
the treatment or prevention of infection or disease caused by
Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
Nucleic acids encoding (I) are used to determine whether a compound binds to
biological sample. (I) is used to determine whether a compound binds to
Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
(I). A composition comprising (I) or a nucleic acid encoding (I), may be
used as a vaccine or diagnostic composition. The disease caused by
Streptococcus that is prevented or treated may be meningitis. Nucleic
acid encoding (I) may be used to recombinantly produce (I) and may be
used in gene therapy. Antibodies to (I) are used for affinity
chromatography, immunoassays, and distinguishing/identifying
Streptococcus proteins
XX
```

SQ	Sequence	1134 BP; 384 A; 200 C; 177 G; 373 T; 0 U; 0 Other;
	Query Match	99.3%; Score 1126; DB 6; Length 1134;
	Best Local Similarity	99.6%; Pred. No. 1.1e-230;
	Matches 1129; Conservative	0; Mismatches 5; Indels 0; Gaps 0;
Qy	1	ATGATTCAAATTTTCAAATAATTCGCACATTAATTATTCATGCTTTAAATACAACCTAAGCGT 60
Db	1	ATGATTCAAATTTTCAAATAATTCGCACATTAATTATTCATGCTTTAAATACAACCTAAGCGT 60
Qy	61	GCTATTAGCACTAAAAATGCCATCCCTATCTTTTCATCAATAAAAATGAAGTCACCTCT 120
Db	61	GCTATTAGCACTAAAAATGCCATCCCTATCTTTTCATCAATAAAAATGAAGTCACCTCT 120
Qy	121	ACAGGAGCTAACCTTTAAACAGGGTCTAACCGGTCAAAATATCAATTGAAAAACACTATTCCCTGTA 180
Db	121	ACAGGAGCTAACCTTTAAACAGGGTCTAACCGGTCAAAATATCAATTGAAAAACACTATTCCCTGTA 180
Qy	181	AGTAATGAAAATGCTGGTTTGCTAAATTAACCTCTCCAGGAGCTATTTTATTAGAAGCTAGT 240
Db	181	AGTAATGAAAATGCTGGTTTGCTAAATTAACCTCTCCAGGAGCTATTTTATTAGAAGCTAGT 240
Qy	241	TTTTTTTATTAATATATTCTCAAGTTTCGCCAGATATTAGTAAATGTTTAAAGAAATTTGAA 300
Db	241	TTTTTTTATTAATATATTCTCAAGTTTCGCCAGATATTAGTAAATGTTTAAAGAAATTTGAA 300
Qy	301	CAACACCAGAGTTGTTTTAAACAGTGGTAAATCAGAGATTACCTTTAAAGGAAAAGATGTT 360
Db	301	CAACACCAGAGTTGTTTTAAACAGTGGTAAATCAGAGATTACCTTTAAAGGAAAAGATGTT 360
Qy	361	GACCAGTATCCTCGTCTACAAGAGTAGTATCAACAGAAAATCCTTTGATTTTAAAAACAAAA 420
Db	361	GACCAGTATCCTCGTCTACAAGAGTAGTATCAACAGAAAATCCTTTGATTTTAAAAACAAAA 420
Qy	421	TTATTGAAGTCTATTATTGCTGAAACAGCTTTTCAGCCAGTTTACAAGAAAAGTCGTCCT 480
Db	421	TTATTGAAGTCTATTATTGCTGAAACAGCTTTTCAGCCAGTTTACAAGAAAAGTCGTCCT 480
Qy	481	ATTTTAAACAGGAGTTCTATTTGTTATTAAGTAATCATAAAGATTTTAAAGCAGTAGCGACT 540
Db	481	ATTTTAAACAGGAGTTCTATTTGTTATTAAGTAATCATAAAGATTTTAAAGCAGTAGCGACT 540
Qy	541	GACTCTCATCGTATGAGCCAAACGTTTAAATCACTTTGGACAATACTTCAGCAGATTTTGATG 600
Db	541	GACTCTCATCGTATGAGCCAAACGTTTAAATCACTTTGGACAATACTTCAGCAGATTTTGATG 600
Qy	601	GTAGTTCTTCCAAGTAAATCTTTGAGAGAAATTTTCAGCAGTATTTACAGATGATTTGAG 660
Db	601	GTAGTTATTTCCAAGTAAATCTTTGAGAGAAATTTTCAGCAGTATTTACAGATGATTTGAG 660
Qy	661	ACGGTTCAGGTAATTTTCTCACCAAGCCAAATCTTTGTTTCAAGAGTGAACATTTCTTTT 720
Db	661	ACGGTTCAGGTAATTTTCTCACCAAGCCAAATCTTTGTTTCAAGAGTGAACATTTCTTTT 720
Qy	721	TATACACGCCCTCTTAGAAGGAAATTAATCCCGATACAGACCGTTTATTAATGACAGAAATTT 780
Db	721	TATACACGCCCTCTTAGAAGGAAATTAATCCCGATACAGACCGTTTATTAATGACAGAAATTT 780
Qy	781	GAGACGGAGGTTGTTTTCAATACCCAAATCCCTTCGCCACGCTATGGAACCGTCCTCTTG 840
Db	781	GAGACGGAGGTTGTTTTCAATACCCAAATCCCTTCGCCACGCTATGGAACCGTCCTCTTG 840
Qy	841	ATTTCTTAATGCTACTCAAAATGGTACTGTTAAAGCTTCAGATTACTCAAAATCATATTTTCA 900
Db	841	ATTTCTTAATGCTACTCAAAATGGTACTGTTAAAGCTTCAGATTACTCAAAATCATATTTTCA 900
Qy	901	GCTCATGTTAACTCACTGAGGTTGGTAAAGTAAACGAGGATTTAGATATTGTTAGTCAG 960
Db	901	GCTCATGTTAACTCACTGAGGTTGGTAAAGTAAACGAGGATTTAGATATTGTTAGTCAG 960
Qy	961	TCGGTAGTGATTTAACTATCAGCTTCAATCCAACTTACCTTATGAGTCTTTAAAAAGCT 1020
Db	961	TCGGTAGTGATTTAACTATCAGCTTCAATCCAACTTACCTTATGAGTCTTTAAAAAGCT 1020

	1021	ATTAAAAGTGAACAGTAATAAATTCATTTCTTATCACAGTTCGACCAATTCACCCCTAAACA	1080
Qy			
Sq			
Df			
	1021	ATTAAAAGTGAACAGTAATAAATTCATTTCTTATCACAGTTCGACCAATTCACCCCTAAACA	1080
Qy	1081	CGAGGGCATGAGGAAGAAGTTTTATCCAAATTAATTACACCAGTAGCAACAAC	1134
Df	1081	CGAGGGCATGAGGAAGAAGTTTTATCCAAATTAATTACACCAGTAGCAACAAC	1134
<b>RESULT 3</b>			
ABK83362			
ID	ABK83362 standard; DNA; 1137 BP.		
XX	AC		
XX	ABK83362;		
DT	12-AUG-2002 (first entry)		
XX	Streptococcus pyogenes dnaN gene.		
DE			
XX	DNA polymerase III holoenzyme subunit protein; replicase; polymerase;		
KW	antibacterial; vaccine; gene; ds.		
KW			
XX	Streptococcus pyogenes.		
OS			
XX	WO200234936-A2.		
FN			
XX	02-MAY-2002.		
PD			
XX	29-OCT-2001; 2001WO-US048396.		
PF			
XX	27-OCT-2000; 2000US-0244023P.		
PR			
XX	(REPL-) REPLIDYNE INC.		
PA			
PI	Mchenry CS, Bullard JM, Janjic N, Manhardt EL, Kery V;		
Pt	Williams JC;		
XX	WPI; 2002-463366/49.		
DR	P-PSDB; ABG61623.		
DR			
XX	Novel bacterial DNA polymerase III subunit proteins and genes encoding		
PT	the proteins, for reconstituting replicases and polymerases for		
PT	sequencing, amplification and screening compounds that modulate		
PT	polymerase.		
XX			
PS	Claim 10; Fig 21A; 268pp; English.		
XX			
CC	The invention describes an isolated bacterial, preferably Streptococcus		
CC	pyogenes DNA polymerase III holoenzyme subunit proteins (I). The protein		
CC	and the polynucleotide encoding it are useful for reconstituting		
CC	replicases and polymerases for sequencing, amplification and screening		
CC	for compounds which modulate the function of polymerase or replace		
CC	antibody to the protein is useful as therapeutic compound to passively		
CC	immunise an animal to protect the animal from bacteria susceptible to		
CC	treatment by such antibodies, preferably Streptococcus pyogenes, as		
CC	reagents in assays to detect infection by such bacteria, and/or as tools		
CC	to screen expression libraries and/or to recover desired proteins from a		
CC	mixture of protein and other contaminants. This sequence encodes a		
CC	Streptococcus pyogenes polymerase III holoenzyme sub-unit		
XX			
SQ	Sequence 1137 BP; 386 A; 200 C; 177 G; 374 T; 0 U; 0 Other;		
<b>Query Match</b> 99.3%; Score 1126; DB 6; Length 1137;			
<b>Best Local Similarity</b> 99.6%; Pred. No. 1.1e-230;			
<b>Matches 1129; Conservative</b> 0; Mismatches 5; Indels 0; Gaps 0;			
Qy	1	ATGATTCAAATTTCAATTAATCGCACAATTATTTATTCATGCTTTAAATACAACCTAACGT	60
Df	1	ATGATTCAAATTTCAATTAATCGCACAATTATTTATTCATGCTTTAAATACAACCTAACGT	60
Qy	61	GCTATTAGCACTAAAAATGCATTTCCTATTCTTTTCATCAATAAAAAATTGAAGTCATCTTCT	120

Db 61 GCTATTAGCACTAAATAATGCCATTCCTATTCTTTCATCAATATAAAATTTGAAGTCACCTTCT 120  
Qy 121 ACAGGAGTAACTTTTAAACAGGGTCTAAACGGTCAATATCAATTTGAAACACATTTCTGTGA 180  
Db 121 ACAGGAGTAACTTTTAAACAGGGTCTAAACGGTCAATATCAATTTGAAACACATTTCTGTGA 180  
Qy 181 AGTAATGAAATGCTGTTGCTAAATACCTCTCCAGGAGCTATTTTATTAGAAGCTAGT 240  
Db 181 AGTAATGAAATGCTGTTGCTAAATACCTCTCCAGGAGCTATTTTATTAGAAGCTAGT 240  
Qy 241 TTTTATTATATATATTTCAAGTTTCCAGATATATAGTATAAATGTTAAAGAAATGAA 300  
Db 241 TTTTATTATATATATTTCAAGTTTCCAGATATATAGTATAAATGTTAAAGAAATGAA 300  
Qy 301 CAACACCAAGTTGTTTAAACAGGTGTAATACAGATTAACCTTAAAGGAAAGATGTT 360  
Db 301 CAACACCAAGTTGTTTAAACAGGTGTAATACAGATTAACCTTAAAGGAAAGATGTT 360  
Qy 361 GACCAGTATCTGCTCTACAGAGTATCAACAGAAATCTTTGATTTTAAACAA 420  
Db 361 GACCAGTATCTGCTCTACAGAGTATCAACAGAAATCTTTGATTTTAAACAA 420  
Qy 421 TTATTGAAGTCTATTTATGCTGTAACAGCTTTTGCAGCCAGTTTACAGAAAGTCTGCT 480  
Db 421 TTATTGAAGTCTATTTATGCTGTAACAGCTTTTGCAGCCAGTTTACAGAAAGTCTGCT 480  
Qy 481 ATTTTAAACAGGAGTTCATATTTGTAATTAAGTAATCAATAAGATTTTAAACAGTAGGACT 540  
Db 481 ATTTTAAACAGGAGTTCATATTTGTAATTAAGTAATCAATAAGATTTTAAACAGTAGGACT 540  
Qy 541 GACTCTCATGATGAGCCAGCTTTTAAATCACTTTTGCAATACACTTTCAGCAGATTTGAT 600  
Db 541 GACTCTCATGATGAGCCAGCTTTTAAATCACTTTTGCAATACACTTTCAGCAGATTTGAT 600  
Qy 601 GTAGTTCTTCCAAAGTAAATCTTTGAGAGATTTTACAGAGTATTTACAGATCATATGAG 660  
Db 601 GTAGTTCTTCCAAAGTAAATCTTTGAGAGATTTTACAGAGTATTTACAGATCATATGAG 660  
Qy 661 ACCGTTGAGTATTTTCTCACCAGCCAAATCTTTGTCAGAGTCAACACATTTCTTTT 720  
Db 661 ACCGTTGAGTATTTTCTCACCAGCCAAATCTTTGTCAGAGTCAACACATTTCTTTT 720  
Qy 721 TATACAGGCTCTTACAGAGAAATATCCGATACAGACCGTTTATTAATGACAGAAATTT 780  
Db 721 TATACAGGCTCTTACAGAGAAATATCCGATACAGACCGTTTATTAATGACAGAAATTT 780  
Qy 781 GAGACGAGGTTGTTTCAATACCAATCCCTTCCGACGCTATGGAACGTCCTTCTTG 840  
Db 781 GAGACGAGGTTGTTTCAATACCAATCCCTTCCGACGCTATGGAACGTCCTTCTTG 840  
Qy 841 ATTTCTAATGCTACTCAAAATGCTACTGTTAAGCTTGATTAATTTCA 900  
Db 841 ATTTCTAATGCTACTCAAAATGCTACTGTTAAGCTTGATTAATTTCA 900  
Qy 901 GCTCATGTTAACTCACTGAGGTTGTTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGG 960  
Db 901 GCTCATGTTAACTCACTGAGGTTGTTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGG 960  
Qy 961 TCTGTAGTATTAATCACTATACGTTTCAATCCACTTACCTTATTTGAGTCTTTTAAAGCT 1020  
Db 961 TCTGTAGTATTAATCACTATACGTTTCAATCCACTTACCTTATTTGAGTCTTTTAAAGCT 1020  
Qy 1021 ATTTAAAGTGAACAGTAAATTTTCTTATCCAGTTCACCACTTCCAGCTTCAAC 1080  
Db 1021 ATTTAAAGTGAACAGTAAATTTTCTTATCCAGTTCACCACTTCCAGCTTCAAC 1080  
Qy 1081 CCAGCGGATGAGGAAAGTATTTATCAATTAATTTACAGGATGAGGAAAC 1134  
Db 1081 CCAGCGGATGAGGAAAGTATTTATCAATTAATTTACAGGATGAGGAAAC 1134

RESULT 4  
ACAS0247

ID ACAS0247 standard; DNA; 1137 BP.

XX ACAS0247;

XX 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #31904.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
drug design; gene.

XX Streptococcus pyogenes.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX P-PSDB; ABU46377.

XX New antisense nucleic acids, useful for identifying proteins or screening

XX for homologous nucleic acids required for cellular proliferation to

XX isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 38117; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

XX the 6213 antisense sequences given in the specification where expression

XX of the nucleic acid inhibits proliferation of a cell. Also included are:

XX (1) a vector comprising a promoter operably linked to the nucleic acid

XX encoding a polypeptide whose expression is inhibited by the antisense

XX nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX polypeptide or its fragment whose expression is inhibited by the

XX antisense nucleic acid; (4) an antibody capable of specifically binding

XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX proliferation or the activity of a gene in an operon required for

XX the gene product or that has an activity against a biological pathway

XX identifying a gene required for cellular proliferation or the biological

XX pathway in which a proliferation-required gene or its gene product lies

XX on a gene on which the test compound that inhibits proliferation of an

XX organism acts; (9) manufacturing an antibiotic; (10) profiling a

XX compound's activity; (11) a culture comprising strains in which the gene

XX product is overexpressed or underexpressed; (12) determining the extent

XX to which each of the strains is present in a culture or collection of

XX strains; or (13) identifying the target of a compound that inhibits the

XX proliferation of an organism. The antisense nucleic acids are useful for

XX identifying proteins or screening for homologous nucleic acids required

XX for cellular proliferation to isolate candidate molecules for rational

XX drug discovery programs, or for screening homologous nucleic acids

XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target

XX prokaryotic essential genes. Note: The sequence data for this patent did

XX not form part of the printed specification, but was obtained in

XX electronic format directly from WIPO at

XX ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 1137 BP; 386 A; 200 C; 177 G; 374 T; 0 U; 0 Other;

Query Match 99.3%; Score 1126; DB 8; Length 1137;  
Best Local Similarity 99.6%; Pred. No. 1.1e-230;  
Matches 1129; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATGATTCAATTTTCAATTAATCGACATTTATTTATTCATGCTTTTAAATACAACTAAAGCT 60  
DB 1 ATGATTCAATTTTCAATTAATCGACATTTATTTATTCATGCTTTTAAATACAACTAAAGCT 60

QY 61 GCTATTAGCACTAAAAATGCCATTCCTATTTCTTTTATCATCAATAAAAAATTTGAAGTCATCTCT 120  
DB 61 GCTATTAGCACTAAAAATGCCATTCCTATTTCTTTTATCATCAATAAAAAATTTGAAGTCATCTCT 120

QY 121 ACAGGAGTAACTTTTAAACAGGCTTAAACGGTCAAAATCAATTTGAAACACATTTCTCTGA 180  
DB 121 ACAGGAGTAACTTTTAAACAGGCTTAAACGGTCAAAATCAATTTGAAACACATTTCTCTGA 180

QY 181 AGTAATGAAATGCTGTTGTTGCTAAATACCTCTCCAGGAGCTATTTTATTAGAAGCTAGT 240  
DB 181 AGTAATGAAATGCTGTTGTTGCTAAATACCTCTCCAGGAGCTATTTTATTAGAAGCTAGT 240

QY 241 TTTTATTAATTAATTTTCAAGTTTGCCAGATATTTAGTATTAATGTTTAAAGAAATGAA 300  
DB 241 TTTTATTAATTAATTTTCAAGTTTGCCAGATATTTAGTATTAATGTTTAAAGAAATGAA 300

QY 301 CAACACCAAGTTGTTTAAACAGGCTTAAACGGTCAAAATCAATTTGAAAGGAAAGATGTT 360  
DB 301 CAACACCAAGTTGTTTAAACAGGCTTAAACGGTCAAAATCAATTTGAAAGGAAAGATGTT 360

QY 361 GACCAAGTATCTCTGCTCAACAGGCTTAAACGGTCAAAATCAATTTGAAAGGAAAGATGTT 420  
DB 361 GACCAAGTATCTCTGCTCAACAGGCTTAAACGGTCAAAATCAATTTGAAAGGAAAGATGTT 420

QY 421 TTAATGAAATCTATTTATTTCAAGTTTGCCAGATATTTAGTATTAATGTTTAAAGAAATGAA 480  
DB 421 TTAATGAAATCTATTTATTTCAAGTTTGCCAGATATTTAGTATTAATGTTTAAAGAAATGAA 480

QY 481 ATTTTAAACAGGAGTTTCAATTTGTTTAAACAGGCTTAAACGGTCAAAATCAATTTGAAAGGAAAGATGTT 540  
DB 481 ATTTTAAACAGGAGTTTCAATTTGTTTAAACAGGCTTAAACGGTCAAAATCAATTTGAAAGGAAAGATGTT 540

QY 541 GACTCTCATGCTATGAGGCAAGTTTAAACAGGCTTAAACGGTCAAAATCAATTTGAAAGGAAAGATGTT 600  
DB 541 GACTCTCATGCTATGAGGCAAGTTTAAACAGGCTTAAACGGTCAAAATCAATTTGAAAGGAAAGATGTT 600

QY 601 GTAGTTCTTCCAGTAATTTTGAAGGAAATTTTGAAGGAAATTTTGAAGGAAATTTTGAAGGAAATTTGAG 660  
DB 601 GTAGTTCTTCCAGTAATTTTGAAGGAAATTTTGAAGGAAATTTTGAAGGAAATTTTGAAGGAAATTTGAG 660

QY 661 ACCGTTGAGTATTTTCTCAGCAAGGCAATTTGTTTGAAGGAAATTTTGAAGGAAATTTTGAAGGAAATTTGAG 720  
DB 661 ACCGTTGAGTATTTTCTCAGCAAGGCAATTTGTTTGAAGGAAATTTTGAAGGAAATTTTGAAGGAAATTTGAG 720

QY 721 TATACAGGCTCTTGAAGGAAATTTTGAAGGAAATTTTGAAGGAAATTTTGAAGGAAATTTTGAAGGAAATTTGAG 780  
DB 721 TATACAGGCTCTTGAAGGAAATTTTGAAGGAAATTTTGAAGGAAATTTTGAAGGAAATTTTGAAGGAAATTTGAG 780

QY 781 GAGACGAGGTTGTTTCAATACCAATCCCTTCCAGGCTTAAACAGGCTTAAACAGGCTTAAACAGGCTTAAACAGGCT 840  
DB 781 GAGACGAGGTTGTTTCAATACCAATCCCTTCCAGGCTTAAACAGGCTTAAACAGGCTTAAACAGGCTTAAACAGGCT 840

QY 841 ATTTTCTAATGCTACTCAAAATGCTACTGTTAGGCTTGAATTTGAGGCTTGAATTTGAGGCTTGAATTTGAGGCTTGA 900  
DB 841 ATTTTCTAATGCTACTCAAAATGCTACTGTTAGGCTTGAATTTGAGGCTTGAATTTGAGGCTTGAATTTGAGGCTTGA 900

QY 901 GCTCATGTTAACTCACTGAGGTTGGTAAGGTTAAACAGGCTTAAACAGGCTTAAACAGGCTTAAACAGGCTTAAACAGGCT 960  
DB 901 GCTCATGTTAACTCACTGAGGTTGGTAAGGTTAAACAGGCTTAAACAGGCTTAAACAGGCTTAAACAGGCTTAAACAGGCT 960

QY 961 TCTGTTAGTATTTAACTATGAGCTTCAATCCAACTTACCTTTTGAAGCTTTTAAAGCT 1020  
DB 961 TCTGTTAGTATTTAACTATGAGCTTCAATCCAACTTACCTTTTGAAGCTTTTAAAGCT 1020

QY 1021 ATTTAAAGTGAACAGTAAAAATTCATTTCTTATCACCAGTTTCGACCATTCACCCATAACA 1080

DB 1021 ATTTAAAGTGAACAGTAAAAATTCATTTTATCACCAGTTTCGACCATTCACCCATAACA 1080

QY 1081 CCAGCGCATGAGGAAGAAAGTTTATCCAAATTAATTAACACGATGACGAACAAAC 1134

DB 1081 CCAGCGCATGAGGAAGAAAGTTTATCCAAATTAATTAACACGATGACGAACAAAC 1134

RESULT 5  
ABK83363  
ID ABK83363 standard; DNA; 1137 BP.  
XX  
AC ABK83363;  
XX  
DT 12-AUG-2002 (first entry)  
XX  
DNA encoding Streptococcus pyogenes DnaN protein.  
DE  
DNA polymerase III holoenzyme subunit protein; replicase; polymerase;  
KW antibacterial; vaccine; gene; ds.  
XX  
Streptococcus pyogenes.  
XX  
WO200234936-A2.  
XX  
02-MAY-2002.  
XX  
29-OCT-2001; 2001WO-US048396.  
XX  
27-OCT-2000; 2000US-0244023P.  
XX  
(REPL-) REPLIDYNE INC.  
XX  
Mchenry CS, Bullard JM, Janjic N, Manhardt BL, Kery V;  
PI Williams JC;  
XX  
WPI; 2002-463366/49.  
DR P-PSDB; ABG61623.  
XX  
Novel bacterial DNA polymerase III subunit proteins and genes encoding  
the proteins, for reconstituting replicases and polymerases for  
sequencing, amplification and screening compounds that modulate  
polymerase.  
XX  
Disclosure; Page 217-218; 268pp; English.  
XX  
The invention describes an isolated bacterial, preferably Streptococcus  
pyogenes DNA polymerase III holoenzyme subunit proteins (I). The protein  
and the polynucleotide encoding it are useful for reconstituting  
replicases and polymerases for sequencing, amplification and screening  
for compounds which modulate the function of polymerase or replicase. An  
antibody to the protein is useful as therapeutic compound to passively  
immunise an animal to protect the animal from bacteria susceptible to  
treatment by such antibodies, preferably Streptococcus pyogenes, as  
reagents in assays to detect infection by such bacteria, and/or as tools  
to screen expression libraries and/or to recover desired proteins from a  
mixture of protein and other contaminants. This sequence encodes a  
Streptococcus pyogenes polymerase III holoenzyme sub-unit  
XX  
SQ Sequence 1137 BP; 387 A; 200 C; 177 G; 373 T; 0 U; 0 Other;

Query Match 99.2%; Score 1124.4; DB 6; Length 1137;  
Best Local Similarity 99.5%; Pred. No. 2.3e-230;  
Matches 1128; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ATGATTCAATTTTCAATTAATCGACATTTATTTATTCATGCTTTTAAATACAACTAAAGCT 60  
DB 1 ATGATTCAATTTTCAATTAATCGACATTTATTTATTCATGCTTTTAAATACAACTAAAGCT 60

QY 61 GCTATTAGCACTAAAAATGCCATTCCTATTTCTTTTATCATCAATAAAAAATTTGAAGTCATCTCT 120  
DB 61 GCTATTAGCACTAAAAATGCCATTCCTATTTCTTTTATCATCAATAAAAAATTTGAAGTCATCTCT 120

RESULT 6  
AAV58285  
ID AAV5  
XX

AC	AAV58285;
XX	
DT	18-DEC-1998 (first entry)
DE	S. pyogenes SP-7-44 genomic DNA.
XX	
KW	Probe; diagnosis; infection; detection; ds.
XX	
OS	Streptococcus pyogenes.
XX	
PN	WO9842845-A1.
XX	
PD	01-OCT-1998.
XX	
PF	23-MAR-1998; 98WO-JP001288.
XX	
PR	25-MAR-1997; 97JP-00071077.
XX	
PA	(FUSO ) FUSO PHARM IND LTD.
XX	
PI	Ueyama H, Abe K, Keshi H, Matsuhisa A;
XX	
DR	WPI; 1998-532009/45.
XX	
PT	New DNA probes, e.g. SP-6-28 or SP-7-44 - useful for, e.g. diagnosis of Streptococcus pyogenes infection.
XX	
PS	Claim 2; Page 19-21; 48pp; Japanese.
XX	
CC	AAV58284-V58289 are novel genomic DNA sequences which can be used as DNA probes for the diagnosis of Streptococcus pyogenes infection. These probes provide for simple and highly specific detection of S. pyogenes in biological samples such as blood
XX	
SQ	Sequence 3200 BP; 1120 A; 547 C; 523 G; 1010 T; 0 U; 0 Other;
	Query Match 74.9%; Score 849; DB 2; Length 3200;
	Best Local Similarity 98.7%; Pred. No. 1.6e-171;
	Matches 866; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
Qy	1 ATGATTCGAATTTCGAATTAATCGCACATTTATTTCATGCTTTTAAATAACAATAAAGCGT 60
Dd	2324 ATGATTCGAATTTTCGAATTAATCGCACATTTATTTCATGCTTTTAAATGCAACTAAACGT 2383
Qy	61 GCATTAGCACTAAAAATGCATTCCCTATTCTTTTCATCAATAAAAAATTGAAGTCACCTCT 120
Dd	2384 GCATTAGCACTAAAATGCCATTCCTATTCTTTTCATCAATAAGATTGAATCACCCTCT 2443
Qy	121 ACAGGAGTAACTTTAACAGGGTCTAACGGTCAAATATCAATGAAAACACTATTCCTGTGA 180
Dd	2444 ACAGGAGTAACTTTAACAGGGTCTAACGGTCAAATATCAATGAAAACACTATTCCTGTGA 2503
Qy	181 AGTAATGAAAATCCTCGTTTGCTAATTACCTCCAGGAGCTATTTATTAGAAGCTAGT 240
Dd	2504 AGTAATGAAAATCCTCGTTTGCTAATTACCTCCAGGAGCTATTTATTAGAAGCTAGT 2563
Qy	241 TTTTTTATTAATATTTATTTCAAGTTTGCCAGATATTAGTATAATGTTTAAAGAAATTTGAA 300
Dd	2564 TTTTTTTATTAATATTTATTTCAAGTTTGCCAGATATTAGTATAATGTTTAAAGAAATTTGAA 2623
Qy	301 CAACACCAAGTTGTTTTTAAACCAAGTGCTAAATCAGAGATTACCTTAAAGGAAAAGATGTT 360
Dd	2624 CAACACCAAGTTGTTTTTAAACCAAGTGCTAAATCAGAGATTACCTTAAAGGAAAAGATGTT 2683
Qy	361 GACCACTATCCTCGTCTACAAGAAGTATCAACAGAAAAATCCCTTTGATTTTTTAAAAACAAA 420
Dd	2684 GACCACTATCCTCGTCTACAAGAAGTATCAACAGAAAAATCCCTTTGATTTTTTAAAAACAAA 2743
Qy	421 TTATTGAAGTCTATTATTCCTGAAACAGCTTTTTCAGCCAGTTTACAAGAAGTCGTCCT 480
Dd	2744 TTATTGAAGTCTATTATTCCTGAAACAGCTTTTTCAGCCAGTTTACAAGAAGTCGTCCT 2803
Qy	481 ATTTTAAACAGGAGTTTCATATTGTATTAAAGTAATCATAAAGATTTTAAAGCAGTAGCGACT 540



Db 2804 ATTTTACAGGAGTTCATATTGTTAAGCAATCAATAAGATTTTAAAGCAGTAGCGACT 2863  
Qy 541 GACTCTCATCGTATGAGCAACGTTTAACTCACTTTGGAC-AATACTTCAGCAGATTTCAT 599  
Db 2864 GACTCTCATCGTATGAGCAACGTTTAACTCACTTTGGACAAATATCTTCAGCAGATTTCGA 2923  
Qy 600 GGTAGTCTTCAAGTAATCTTTGAGAGAAATTTTCAGCAGATTTCAGCAGATTTCGA 659  
Db 2924 TGTGGTTATTCGAAGTAAATCTTTTCAGAGAAATTTTCAGCAGATTTCAGCAGATTTCGA 2983  
Qy 660 GACCGTTCAGGATTTTCTCACCAGCAATCTTGTTCAGAGTGAACATTTCTTT 719  
Db 2984 GACCGTTCAGGATTTTCTCACCAGCAATCTTGTTCAGAGTGAACATTTCTTT 3043  
Qy 720 TTATACAGCGCTCTTAGAGGAAATATTCGCCATACAGACCGTTTATTAAATGACAGAAAT 779  
Db 3044 TTATACAGCGCTCTTAGAGGAAATATTCGCCATACAGATCGTTTATTAAATGACAGAAAT 3103  
Qy 780 TGAGACGAGGTTGTTTCAATACCAATCCCTTCGCCACGCTATGGAAACGTCCTTCTT 839  
Db 3104 TGAGACGAGGTTGTTTCAATACCAATCCCTTCGCCACGCTATGGAAACGTCCTTCTT 3163  
Qy 840 GATTCTTAATGCTACTCAAAATGGTACTGTTAAGCTT 876  
Db 3164 GATTCTTAATGCTACTCAAAATGGTACTGTTAAGCTT 3200

RESULT 7  
ACA47964  
ID ACA47964 standard; DNA; 1134 BP.  
XX ACA47964;  
XX  
XX 19-JUN-2003 (first entry)  
XX Prokaryotic essential gene #29621.  
XX  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
XX drug design; gene.  
XX Streptococcus mutans.  
XX  
XX WO200277183-A2.  
XX  
XX 03-OCT-2002.  
XX  
XX 21-MAR-2002; 2002WO-US009107.  
XX  
XX 21-MAR-2001; 2001US-00815242.  
XX 06-SEP-2001; 2001US-00948993.  
XX 25-OCT-2001; 2001US-0342923P.  
XX 08-FEB-2002; 2002US-00072851.  
XX 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
XX P-PSDB; ABU44094.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX PT for homologous nucleic acids required for cellular proliferation to  
XX PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 14; SEQ ID NO 35834; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX CC the 6213 antisense sequences given in the specification where expression  
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 1134 BP; 380 A; 194 C; 179 G; 381 T; 0 U; 0 Other;

Query Match 60.9%; Score 690.8; DB 8; Length 1134;  
Best Local Similarity 75.6%; Pred. No. 7.9e-138;  
Matches 857; Conservative 0; Mismatches 277; Indels 0; Gaps 0;  
Qy 1 ATGATTCAATTTCAATTAATCGCACATTATTATTCATGCTTTTAAATACACTAACT 60  
Db 1 ATGATAAATTTTCAATTAATAAGTTTTTCTTACAGCCTTAATATGCTTACCAGCGA 60  
Qy 61 GCTATTAGCACTAAAAATGCCATTCCCTATTCTTTTCATCAATAAAAAATGAAGTCACTTCT 120  
Db 61 GCTATTAGTCTAAAAATGCTATTCTTCTTCTAGTCTTTAAAAATTTGAAGTGAATCT 120  
Qy 121 ACAGAGTAACTTTTAAACAGGGTCTAACGGTCAATATCAATTAATGAAAAACACTATTCTGTGA 180  
Db 121 CAGTCCATTACTTTAAACAGGGCTCTAATGAGCAAAATTTCTATTGAAAAATACGATTCAGCT 180  
Qy 181 AGTAATGAATGCTGTGTTGCTTAATACCTCTCCAGGAGCTATTTTATTAGAACTAGT 240  
Db 181 GAAGAAGAAATGCTGGACTATTGGTTACTTCTTCAGGAGCTATCTTACTTGAAGCTAAT 240  
Qy 241 TTTTATTATAATATTATTCAAGTTTGGCAGATATTAGTATAAAATGTTAAAGAAAATTGAA 300  
Db 241 TTTTATTATAATATTATTCAAGCTCTACCTGATATCACTTTAGATTTTGAAGAAATGAA 300  
Qy 301 CAACCAAGTTGTTTAAACCAAGTGGTAAATCAGAGATTACCTTAAAGGAAAAAGATGTT 360  
Db 301 CAACATCAAGTTGTTTAAATAGTGGAAAAATCAGAAATTTACACTTTAAAGGTAAGATGTG 360  
Qy 361 GACAGTATCTCGTCTACAGAGAGTATCAACAGAAATCCCTTGAATTTTAAAGCAAAA 420  
Db 361 GAACAAATATCTCGTTTCAGAGAGTGGGAAACAAATAATCTTTTAAATTTTAAAGCAAAA 420  
Qy 421 TTATTGAAGTCTATTATTGCTGAAAACAGCTTTTTCAGGCGAGTTTACAAAGAAAGTCTCT 480  
Db 421 TTGTTAAAAACAATATTCTGAAACAGCTTTTTCGCTAGTACCCAGAAAGTCAACCA 480  
Qy 481 ATTTTAAACAGAGTTTCATATTGTTAATAGTAAATCAATAAGATTTTAAAGCAGTAGCGACT 540  
Db 481 ATTTTAAACAGGTTTCATTGTTGTTATGCAAAATCATAAAGAGTTTAAAGCTGTTGCTACA 540  
Qy 541 GACTCTCATCTATGAGCAACGTTTAACTCACTTTGGCAATATCTTACGACAGATTGATG 600  
Db 541 GACTCTCATCGATGAGTCAAGAAAAATTAACCTCTTGACCATTCATCATGATGATTGAT 600



QY 661 ACCGTTGAGGTAATTTTCTCACCAGCCAAATCTTGTTCAGAAGTGAACATTTCTTTT 720  
 Db 661 ACTGTAGAGGTTTCTTCTCAGATAGTCAAAATGTTTATTAGAAAGTGAATAATCAGCTTC 720  
 QY 721 TATACACGCTCTTTAGAAAGGAAATATCCGATACAGACCGTTTATTATGACAGAAATTT 780  
 Db 721 TATACACGCTCTTTAGAAAGGAAATATCCGATACAGACCGTTTATTATGACAGAAATTT 780  
 QY 781 GAGACGAGGTTGTTTCAATACCCAACTCCCTTCGCCACGCTATGGAACGCTCTTCTTG 840  
 Db 781 GAAACCGAAATATCTTTTAATACAAATGCTTTACGCCATGCTATGGAACGCTCTTATTTA 840  
 QY 841 ATTTCTAATGCTACTCAAAATGTTTAAAGTCTTGAAGTCTTGAGATTAATCTCAAAATCAATTTCA 900  
 Db 841 ATTTCTAATGCTACTCAAAATGTTTAAAGTCTTGAAGTCTTGAAGTCTTGAAGTCTTCA 900  
 QY 901 GCTCATGTTAACTCACTCAGTCTGTAAGTCTGTAAGTCTGTAAGTCTGTAAGTCTGTAAGTCTG 960  
 Db 901 GCTCATGTTAACTCACTCAGTCTGTAAGTCTGTAAGTCTGTAAGTCTGTAAGTCTGTAAGTCTG 960  
 QY 961 TCTGTAGTGAATTTAACTATCAGCTTCAATCCAACTTACCTTATTTAGTCTTTTAAAGCT 1020  
 Db 961 AAAGGTGATAGTTTAAATATTAGCTTTTAACTCACTTACCTTAAATGAACTTTTAAAGCA 1020  
 QY 1021 ATTTAAAGTGAACAGTAAATTTCAATTTCTTTATCACCAGTTCGACCAATTCACCCCTAACA 1080  
 Db 1021 GTAAAAAGCAACAGTACGATTCGATTTTATTTCTCCAGTACGTCCTTACTTTTGACA 1080  
 QY 1081 CCAGCGATGAGGAAGTAAATTTTATCCAAATTAATTTACACGATGACGAAC 1134  
 Db 1081 CCTGTGAGGATGACTGAAGATTTTATCCAAATTAATTTTCTGTTCTGTTACTAAC 1134

RESULT 9

ABN71527\_21  
 Continuation (22 of 22) of ABN71527 from base 2100001 (Streptococcus polynucleotide SEQ  
 WP Sequence split into 22 fragments LOCUS ABN71527 Accession ABN71527

Fragment Name	Begin	End
WP ABN71527_00	1	110000
WP ABN71527_01	100001	210000
WP ABN71527_02	200001	310000
WP ABN71527_03	300001	410000
WP ABN71527_04	400001	510000
WP ABN71527_05	500001	610000
WP ABN71527_06	600001	710000
WP ABN71527_07	700001	810000
WP ABN71527_08	800001	910000
WP ABN71527_09	900001	1010000
WP ABN71527_10	10000001	1110000
WP ABN71527_11	11000001	1210000
WP ABN71527_12	12000001	1310000
WP ABN71527_13	13000001	1410000
WP ABN71527_14	14000001	1510000
WP ABN71527_15	15000001	1610000
WP ABN71527_16	16000001	1710000
WP ABN71527_17	17000001	1810000
WP ABN71527_18	18000001	1910000
WP ABN71527_19	19000001	2010000
WP ABN71527_20	20000001	2110000
WP ABN71527_21	21000001	2155561

Query Match 56.5%; Score 641.2; DB 6; Length 55561;  
 Best Local Similarity 72.8%; Pred. No. 5.9e-127;  
 Matches 826; Conservative 0; Mismatches 308; Indels 0; Gaps 0;

QY 1 ATGATTCATTTTCAATTAATCGCATTTATTTATTCATGCTTTTAAATACAACTAAACGT 60  
 Db 28392 ATGATTCATTTTCAATTAATTAATAAATTTCTTCTGCTGCTCTTAACGCTAACCAACGA 28451  
 QY 61 GCTATTAGCACTAAATGCGCATTCCTTCTTCTCATCAATAAATAATTAAGTCACTTCT 120  
 Db 28452 GCTATTAGTCATAAATAATGCGCATTCCTTCTTCTCATCAATAAATAATTAAGTCACTAGG 28511

RESULT 10  
 ABX08064  
 ID ABX08064 standard; DNA; 1134 BP.

QY 121 ACAGAGTAACCTTTTAAACAGGGTCTAACGGTCAAAATATCAATTTGAAAAACATTTCTCTGA 180  
 Db 28512 GATGCTATCTTTTAAACGGGGTCAAAATGGAACAAATTTCAATTTGAAAAATCTATTCTCTGCT 28571  
 QY 181 AGTAATGAATGCTGGTTTCTTAATACCTCTCCAGAGCTATTTTATTATGAGAGCTAGT 240  
 Db 28572 TCAAAATGAATGAGGTTTACTAGTAACGAATCCAGGCTCTATTGTTGTAAGAGCTGGT 28631  
 QY 241 TTTTATTATTAATATTATTCAAGTTTGGCAGATATTAGTATAAAATGTTTAAAGAAATGAA 300  
 Db 28632 TTCTTTATTATTAATATTATTCAAGTTTACCAGATGTAACCTTTAGAAATTTACAGAGATTGAA 28691  
 QY 301 CAACACCAAGTTGTTTAAACAGTGGTAAATCAGAGATTAACCTTTAAAGGAAAAAGATGTT 360  
 Db 28692 CAACATCAAAATGTTTCTTACTAGTGGAAAAATCAGAGATTACTTTGAAAGGTAAGGATGTC 28751  
 QY 361 GACAGATATCTCTGCTCTACAGAAAGTATCAACAGAAATCTCTTGTATTTTAAACAAAA 420  
 Db 28752 GATCAATACCTCTGCTGAGGAAATGACACAGATCTCCATTAACATTTAGAAATGAA 28811  
 QY 421 TTATTGAAGTCTATTATTGCTGAAACAGCTTTTGCAGCCAGTTTACAAAGAAAGTCTGCT 480  
 Db 28812 CTGTTAAATCAATTTATTAAAGTCTGCTTTGCTGCTAGCCAAAGAACGCTCCA 28871  
 QY 481 ATTTTAAACAGAGTTCATATTGTAATTAAGTAATCATAAAGATTTTAAAGCAGTAGCGACT 540  
 Db 28872 ATCTTAAACAGGTTTCATTTGTTGTTATCAGTCAAAATAAATACTTTTAAAGGCTGTTGCGACA 28931  
 QY 541 GACTCTCATCTGATGAGCCCAAGTTTAACTACTTTGGCAATATCTTACAGAGATTTGATG 600  
 Db 28932 GATTACACACCGTATGAGTCAACGCACTTTTCCAAATAGAGAAATCGGCTAAATTAATTTGAT 28991  
 QY 601 GTAGTTCTTCAAGTAAATCTTTGAGAGAAATTTTACAGAGATTTTACAGATGATATTGAG 660  
 Db 28992 TTGTTGTTTCAAGTAAATCTTTCGAGAAATTTTTCGGCTGTTTTCAGAGATGATTTGAA 29051  
 QY 661 ACCGTTGAGGTAATTTTCTTCAACCAAGCAAAATCTTGTTCAGAAGTGAACATTTCTTTT 720  
 Db 29052 ACTGTAGAGGTTTCTTCTCAGATAGTCAAAATGTTTATTAGAAAGTGAATAATCAGCTTC 29111  
 QY 721 TATACAGCCTCTTTAGAGGAAATTTTCCCGATCAGACAGCCGTTTATTATGACAGAAATTT 780  
 Db 29112 TATACAGCCTTTGCTTGAAGGAAATCTACCTGATCTGATCGCTCTTAACTAAATCAGTTT 29171  
 QY 781 GAGACGAGGTTGTTTCTCAATACCAATCCCTTCCGACGCTATGGAACGCTGCTTTCTTG 840  
 Db 29172 GAAACCGAAATTAATCTTTAATACAAATGCTTTACGCCATGCTATGGAACGCTTTATTTA 29231  
 QY 841 ATTTCTAATGCTACTCAAAATGTTGTAAGCTTTGAGATTAATCTCAAAATCATATTTCATCA 900  
 Db 29232 ATTTCTAATGCTACTCAAAATGTTGTAAGCTTTGAGATTAATCTCAAAATCATATTTCATCA 900  
 QY 901 GCTCATGTTAACTCAGCTGAGGTTGTTGTAAGCTTAACGAGGATTTAGATATTGTTAGTCTAG 960  
 Db 29292 GCTCATGTTAACTCAGCTGAGGTTGTTGTAAGCTTAACGAGGATTTAGATATTGTTAGTCTAGTCTT 29351  
 QY 961 TCTGTAGTGAATTTAACTATCAGCTTCAATCCAACTTACCTTATTGAGTCTTTTAAAGCT 1020  
 Db 29352 AAAGGTGATAGTTTAAATATTAGCTTTTAACTCACTTAACTTAAATGTAATTTTAAAGCA 29411  
 QY 1021 ATTTAAAGTGAACAGTAAATTTCAATTTCTTTATCACCAGTTCGACCAATTCACCCCTAACA 1080  
 Db 29412 GTAAAAAGCAACAGATTACGATTCGATTTATTTCTCCAGTACGCTCCATTTACTTTTGACA 29471  
 QY 1081 CCAGCGATGAGGAAGAAAGTTTATCAATTAATTTATACACGATGACGAAC 1134  
 Db 29472 CCTGTGAGGATGACTGAAGATTTTATCCAAATTAATAATCTCTGTTCTGTTACTAAC 29525

XX AC	ABX08064;	QY	1	ATGATTCAATTTTCAATTAATGCGCATTATTTATTCATGCTTTAAATACAACTAAAGT	60
XX DT	27-OCT-2003 (revised)	Db	1		
XX DT	11-FEB-2003 (first entry)	Db			
XX DE					
XX DE	S. pneumoniae type 4 strain coding region #2352.	QY	61	GCTATTAGCAGCTAAAGATGCGATTCCTATTCTTTTCATCAATAAATAATGAAGTCACTTCT	120
XX KW	Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;	Db	61		
XX KW	ear infection; antinflamatory; antibacterial; immunostimulant;	QY	121	ACAGGAGTAACCTTTAAACAGGGTCTAACCGTCAATATCAATTTGAAACACATTTCTCTGA	180
XX KW	auditory; respiratory; gene therapy; vaccine.	Db	121		
XX OS		QY	181	AGTAATGAAGATGCTGGTTTCTTAATCTCTCCAGAGCTATTTTATTAAGAGTAGT	240
XX PN	Streptococcus pneumoniae; type 4 strain.	Db	181		
XX PN	WO200277021-A2.	QY	241	TTTTTATTAATATTATTTCAAGTTTCCAGATTTAGTATTAATGTTAAAGAAATTGAA	300
XX PD	03-OCT-2002.	Db	241		
XX PD		QY	301	CAACCAAGTTGTTTAAACAGTGGTAAATCAGAGATTACCTTTAAAGGAAAAGATGTT	360
XX PF	27-MAR-2002; 2002WO-1B002163.	Db	301		
XX PR	27-MAR-2001; 2001GB-00007658.	QY	361	GACCAGTATCTCGTCTACAAGAGTATCAACAGAAATCTTTGATTTTAAACAAAA	420
XX PR	(CHIR-) CHIRON SPA.	Db	361		
XX PA	(GENO-) INST GENOMIC RES.	QY	421	TTATTGAAGTCTATTATTGCTGAAACAGCTTTTGACCGAGTTTACAGAAAGTCGTCCT	480
XX PI	Masignani V, Tettelin H, Fraser C;	Db	421		
XX PI	WPI; 2003-040579/03.	QY	481	ATTTTAAACAGAGTTCTATATTGATTAAAGTAATCATAAAGATTTTAAAGCAGTAGGACT	540
XX DR	P-PSDB; ABU02772.	Db	481		
XX DR		QY	541	GACTCTCATGATGAGCCAAAGCTTTTAAATCACTTTGACCAATTTTCAGCAGATTTGATG	600
XX PT	New proteins and nucleic acid molecules from Streptococcus pneumoniae,	Db	541		
XX PT	useful as medicaments for treating or preventing a disease or infection	QY	601	GTAGTTCTTCCAAAGTAAATCTTTGAGAGAAATTTTACAGAGTATTTACAGATGATTTAG	660
XX PT	due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or	Db	601		
XX PT	ear infection.	QY	661	ACCGTTGAGGTATTTTCTCACCAAGCCAAATCTTTGAGAGTGAACACATTTCTTTT	720
XX PS	Claim 6; SEQ ID NO 4703; 56pp; English.	Db	661		
XX PS		QY	721	TATACACGCTCTTAGAAGGAAATTTCCGATACAGACCGTTTATTAATGACAGATTT	780
XX CC	The invention relates to a protein comprising or having at least 50%	Db	721		
XX CC	identity to any of the 2469 amino acid sequences, identified in the	QY	781	GAGACGAGGTGTTTTTCAATACCCCAATCCCTTCCGACGCTATGGAACGTCCTTTCTTG	840
XX CC	specification (available on a computer readable format), or its fragment,	Db	781		
XX CC	expressed from 2469 of 2489 identified DNA coding regions from the	QY	841	ATTTCTAATGCTACTCAAAATGGTACTGTTAGCTTGAGATTACTCAAAATCATATTCA	900
XX CC	Streptococcus pneumoniae type 4 strain genomic sequence appearing as	Db	841		
XX CC	AB556454. Also included are an antibody which binds one of the proteins,	QY	901	GCTCATGTTTAACTCACCTGAGGTGGTAAAGTAAACAGAGATTAGATATTGTTAGTCAG	960
XX CC	treating a patient by administering the protein, DNA or antibody (in a	Db	901		
XX CC	composition), a kit comprising first and second primers, which are the	QY	961	TCGTGTAGTGAATTAACACTATGAGTTCAATCAACTTACCTTATGAGTCTTTAAAGCT	1020
XX CC	nucleic acid cited above or fragments between nucleotides 8-100 of a	Db	961		
XX CC	sequence not defined in the specification, for amplifying a target	QY	1021	ATTAAGGTGAAACAGTAAAGAAATTCATTTCTTATCACCGATTCGACCTTACACCTTAACA	1080
XX CC	the first primer is substantially complementary to the target sequence	Db	1021		
XX CC	and the second primer is substantially complementary to the target sequence	QY			
XX CC	of the target sequence, and where the parts of the primers having	Db			
XX CC	substantial complementarity define the termini of the target sequence to	QY			
XX CC	be amplified, assay comprising contacting a test compound with the	Db			
XX CC	protein, and determining whether the test compound binds to the protein	QY			
XX CC	and a Streptococcus pneumoniae bacterium, where one or more genes	Db			
XX CC	encoding the proteins has been rendered inactive. The proteins, nucleic	QY			
XX CC	acid molecules, antibody and compositions are useful as medicaments for	Db			
XX CC	treating or preventing a disease or infection due to streptococcus	QY			
XX CC	bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis	Db			
XX CC	media or ear infection. They are also useful in developing vaccines,	QY			
XX CC	diagnostics and antibiotics. The methods are useful for identifying	Db			
XX CC	immunodominant proteins. The present sequence is one of the 2489	QY			
XX CC	identified coding region from the genomic sequence. Note: The sequence	Db			
XX CC	data for this patent did not form part of the printed specification, but	QY			
XX CC	was obtained in electronic format directly from WIPO at	Db			
XX CC	ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to	QY			
XX CC	standardise OS field)	Db			
XX SQ	Sequence 1134 BP; 384 A; 208 C; 182 G; 360 T; 0 U; 0 Other;				

Query Match 53.6%; Score 608.2; DB 10; Length 1134;  
 Best Local Similarity 71.1%; Pred. No. 3.3e-120;  
 Matches 805; Conservative 0; Mismatches 328; Indels 0; Gaps 0;

```
QY 1081 CCAGCGATGAGGAGGAGTTTATCCAAATTAATACACGATGACGACAAA 1133
Db 1081 CCAGCAGATCTGACGAGAGCTTCAGCGCTCATACACGATTCGTAACAAA 1133

RESULT 11
AAV52153
ID AAV52153 standard; DNA; 21338 BP.
XX
AC AAV52153;
DT 23-OCT-1998 (first entry)
DE Streptococcus pneumoniae genome fragment SEQ ID NO:20.
XX
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS Streptococcus pneumoniae.
XX
FN WO9818931-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US019588.
XX
PR 31-OCT-1996; 96US-0029960P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;
PI Dougherty BA;
XX
DR WPI; 1998-272225/24.
XX
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus pneumoniae.
XX
PS Claim 1; Page 257-269; 1409pp; English.
XX
SC The present invention describes a computer readable medium which has the
CC nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S. pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1 to
CC 391, identifying members of the library which contain sequences that
CC hybridise to the target sequence and isolating the nucleic acid molecules
CC from the members; or (b) isolating mRNA, DNA or cDNA produced from an
CC organism, amplifying nucleic acid molecules whose nucleotide sequence is
CC homologous to amplification primers derived from the fragment of the S.
CC pneumoniae genome to prime the amplification and isolating the amplified
CC sequences. The computer readable medium can be used in a computer-based
CC system for identifying fragments of the S. pneumoniae genome of
CC commercial importance, or expression modulating fragments of the S.
CC pneumoniae genome. Products from the present invention can be used in
CC diagnosis kits and assays, and pharmaceutical compositions and vaccines
CC for S. pneumoniae
XX
SQ Sequence 21338 BP; 7339 A; 3446 C; 4112 G; 6441 T; 0 U; 0 Other;
Query Match 53.6%; Score 608.2; DB 2; Length 21338;
Best Local Similarity 71.1%; Pred. No. 5.5e-120;
Matches 805; Conservative 0; Mismatches 328; Indels 0; Gaps 0;
QY 1 ATGATTCATTTTCAATTAATCGCACATTTATTTATTCATGCTTTAAATACAACTAAACGCT 60
Db 7076 ATGATTCATTTTCAATTAATAAATAATTTATTTCTACAAGCATTAAATACTACTAAGAGA 7135
```

```
QY 61 GCTATTAGCACTAAAAATGCCATTCTTATCTTTTCATCAATAAAATTAAGTCACTTCT 120
Db 7136 GCTATTAGTCTAAAAATGCCATTCTTATTTATCAACAGTAAAAATTAAGGTGACCAAT 7195
QY 121 ACAGGAGTAACTTTAAACAGGGTCTAAACGGTCAATATCAATTAATGAAAAACACTATTCTCTGA 180
Db 7196 GAAGGTATTACTTTAAATGGTTCAATATGGTCAAAATTTCAATTTGAAAAATTTTATTTCTCAA 7255
QY 181 AGTAATGAAAAATGCTGGTTTCTTAATACCTCTCCAGGAGCTATTATTTATAGAACCTAGT 240
Db 7256 AAAAATGAAGATGCTGGTTTCTTAATTAATTTAGGTTCGATCCCTTCTTGAAGCTTCT 7315
QY 241 TTTTATTAATAATATTTTCAAGTTTGCCAGATATTAGTATATAATATGTTTAAAGAAATGAA 300
Db 7316 TTTTATTAATATGATATCTAGTTTACCTGATGTAACCTTGTATTTTAAAGAAATGAA 7375
QY 301 CAACACAAAGTTGTTTAAACAGGTGTTAAATCAGAGATTACCTTAAAGGAAAGAGATGTT 360
Db 7376 CAAATCAAAATGTTTAAACAGGTGTTAAATCAGAAATTTACCTTAAAGGAAAGAGATGAG 7435
QY 361 GACCAATATCTCGTCTACAAAGAGATATCAACAGAAAAATCCCTTTGATTTTAAAAACAAA 420
Db 7436 GAACAAATATCCAGAAATCCAGAAATTTTCAAGAGCACTCTCTTAAATCTTGAACAAATA 7495
QY 421 TTATTTGAAGTCTATTTTCTGAAACAGCTTTTGCAGCCAGTTTACAGAAAGTCTGCTCT 480
Db 7496 TTACTCAAGAAATTTTAAATGAAACAGCTTTTCTGCAAGTACACAAAGAGAGTCTGCTCG 7555
QY 481 ATTTTAAACAGGATTCATATTTAGTATTAATAGTAATCATAAAGATTTTAAAGCAGTACGACT 540
Db 7556 ATTTTAAACAGGTGCTCACTTCTGATTTAGTCAACAAAGAGTTTAAACAGTTCGCAACA 7615
QY 541 GACTCTCATCTGATGAGCCAAAGTTTAAATCACTTTGGAACAATCTTCAAGCAGATTTGATG 600
Db 7616 GACTCTCATCTGATGAGCCAAAGTTTAAATCACTTTGGAACAATCTTCAAGCAGATTTGATG 7675
QY 601 GTAGTTCTTCAAGTAATCTTTGAGAGAAATTTTCAAGCAGTATTTACAGATGATATTGAG 660
Db 7676 GTCTGTAATTCCTAGCGCTTCTTACGCGAAATTTTCAAGCAGTATTTACAGATGATATTGAG 7735
QY 661 ACCGTGAGGATTTTCTTCAACAGCCAAATCTTGTTCAGAGTGACACATTTCTTTT 720
Db 7736 ACTGTAGAGATTTTCTTGTGCAATTAACCAATCTCTTTAGAGGAAATTAATAGCTTTC 7795
QY 721 TATACAGCCCTCTTAGAGGAAATTTATCCGATACAGACCGTTTATTAATGACAGAAATTT 780
Db 7796 TATCTGCTCTCTAGAGGAAATTTATCTCTGATACAGATCGCTTGTATTTCCACAGACTTT 7855
QY 781 GAGACGGAGTTGTTTCAATACCAATCTCTTCCGACGCTATGGAACGTCGCTCTTTG 840
Db 7856 AACACTACTATTACTTTTAAATGTTGTAATTTACGCCAGTCAATGAGCGTGCCTGCTTT 7915
QY 841 ATTTCTAATGCTACTCAAAATGTTGTTAGCTTGAATTTACTTCAAAATCATATTTCA 900
Db 7916 TTATCAAGTGGACTCAAAATGTTGTTAGCTTGAATTTAGAGTGGGGTGTGTAGC 7975
QY 901 GCTCATGTTAACTCACTGAGTTGGTAAAGTAAACGAGGATTTAGATATTGTTAGTCAG 960
Db 7976 GCCCATGTTCACTCTCCAGAGTTGGTAAAGTAAACGAGGAAATCGATATGATCAGTT 8035
QY 961 TCTGGTAGTATTAACTATATCAGCTTCAATCCAACTTACCTTATTTAGTCTTTTAAAGCT 1020
Db 8036 ACTGGTAGAATTTGACCATTAGTTTCAACCCAACTTACTTGTGATTTCTTAAAGCT 8095
QY 1021 ATTTAAAGTGAACAGTAAAAATTTCTTATCACCAGTTCGACCAATTCACCCCTAACA 1080
Db 8096 TTAATAGCAAGAAAGGTGACTATTAGCTTTTATCTCAGCTGTTGCTGCTTACTTTCTTGTG 8155
QY 1081 CCAGCGATGAGGAGGAAAGTTTATCCAAATTAATTAACACGATGACGACAAA 1133
Db 8156 CCAGCAGATCTGAGGAGAGCTTCAATGAGCTCATACACGATTCGTAACAAA 8208
```

RESULT 12  
ABS56454\_20  
Continuation (21 of 22) of ABS56454 from base 2000001 (Streptococcus pneumoniae type 4)  
WP Sequence split into 22 fragments LOCUS ABS56454 Accession AB56454

Fragment Name	Begin	End
ABS56454_00	1	110000
ABS56454_01	100001	210000
ABS56454_02	200001	310000
ABS56454_03	300001	410000
ABS56454_04	400001	510000
ABS56454_05	500001	610000
ABS56454_06	600001	710000
ABS56454_07	700001	810000
ABS56454_08	800001	910000
ABS56454_09	900001	1010000
ABS56454_10	1000001	1110000
ABS56454_11	1100001	1210000
ABS56454_12	1200001	1310000
ABS56454_13	1300001	1410000
ABS56454_14	1400001	1510000
ABS56454_15	1500001	1610000
ABS56454_16	1600001	1710000
ABS56454_17	1700001	1810000
ABS56454_18	1800001	1910000
ABS56454_19	1900001	2010000
ABS56454_20	2000001	2110000
ABS56454_21	2100001	2162598

Query Match 53.6%; Score 608.2; DB 10; Length 110000;  
Best Local Similarity 71.1%; Pred. No. 7.3e-120;  
Matches 805; Conservative 0; Mismatches 328; Indels 0; Gaps 0;

QY	1	ATGATTCAATTTTCAATTAATCGCATTTATTTTCATCGTTTAATAACAACCTAAACGT	60
Db	56227	ATGATTCAATTTTCAATTAATAAATAATTTATTTCTACAGCATTAAATGACTACGAGA	56286
QY	61	GCTATTAGCACTAAAAATGCCATTCCTATTCTTTTCATCAATAAAAAATGAAGTCACCTCT	120
Db	56287	GCTATTAGTCTTAAAATGCCATTCCTATTATTCACAGCTAAAATGGAGTGCCAAAT	56346
QY	121	ACAGGAGTAATTTAACAGGGTCTAACGGTCAAATATCAATGAAGAAACATATTCCTGTA	180
Db	56347	GAAGGTATTACTTTAATTTGGTTCAATTTGTTCAATTTCAATTTGAAAATTTATTTCTCAA	56406
QY	181	AGTAATCAAAATGCTGGTTTGCTAATFACCTCTCCAGGAGCTATTTTATTAGAAGCTAGT	240
Db	56407	AAAATGAAGATGCTGGTTTGTTAACTTTCTTTAGGTTTCAATCTTTGAGCTTCT	56466
QY	241	TTTTTTTAAATATTTTCAAGTTTGCCAGATATTAGTATAATGTTAAAGAAATGAA	300
Db	56467	TTCTTTATCAATGTAGTATCTAGTTTACCCTGATGTAATCTCTGATTTTAAAGAAATGAA	56526
QY	301	CACACCAAGTGTGTTTAAACAGTGGTAAATCAGAGTTACCTTAAAGGAAAAGATGTT	360
Db	56527	CAAAATCAAAATGTTTTTAAACAGTGGCAATCAGAAATACCCCTAAAGGAAAAGATGAGC	56586
QY	361	GACCAGTATCCTCGTCTACAGAAGTATCAACAGAAAATCCTTTGATTTTAAAAACA AAA	420
Db	56587	GAACAATATCCAGAAATCCAAGAAATTTACAGAAGCACTCCTTTTAACTTTGAACA AAA	56646
QY	421	TTATTGAAGTCTATTATTTGCTGAAACAGCTTTTGCAGGCAAGTTTACAGAAAGTCGTCCT	480
Db	56647	TTACTCAAGAAAATTTAATTAATGAACAGCCCTTTGCTGCAAGTACACAAGAGATCGTCCG	56706
QY	481	ATTTTAAACAGGAGTTCATATTGTTATTAAAGTAATCAVAAAGATTTTAAAGCAGTAGCGACT	540
Db	56707	ATTTTAAACAGGTGTCACCTTCGTTATTCAGTCAACAAGAGTTTAAAAACAGTTGCAACA	56766
QY	541	GACTCTCATGTTAGGCCAAAGTTTAAATCACTTTGGA CAATCTTCAGCAGATTTTGATG	600
Db	56767	GACTCTCATGCGCTAAGCCAGAAAAAATTTGACTCTTGAAAAAATAATGATGATTTGAT	56826







Claim 14; SEQ ID NO 37498; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway; (8) required for proliferation, or that inhibits cellular proliferation; (9) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

Seq. Sequence 1137 BP; 384 A; 210 C; 182 G; 361 T; 0 U; 0 Other;

Query Match 53.2%; Score 603.4; DB 8; Length 1137;  
 Best Local Similarity 70.8%; Pred. No. 3.5e-119;  
 Matches 802; Conservative 0; Mismatches 331; Indels 0; Gaps 0;

QY 1 ATGATTCATTTTCAATTAATCGCACATTATTATTCATGCTTTAAATACAACTAAAGCT 60  
 DB 1 ATGATTCATTTTCAATTAATTAATAAATTTATTTCTACAGCAATTAATTTACTAAGAGA 60

QY 61 GCTATTAGCACTAAATATGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 120  
 DB 61 GCTATTAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 120

QY 121 ACAGAGTAACTTTTAAACAGGCTTCAACGCTCAATATCAATTTGAAACACATTTCTCTGTA 180  
 DB 121 GAAGGTGTTACTTTTAAATGGTTCAAATGGTCAAAATTTTCAATTTGAAATTTTATTTCTCAA 180

QY 181 AGTAATGAAATGCTGGTGTGCTTAAATACCTCTCCAGGAGCTATTTTATTAGAGAGTCTAGT 240  
 DB 181 AAAAATGAAGATGCTGGTGTGCTTAAATACCTCTCTTAGGTTGATCTTCTTCTTCTTCTTCT 240

QY 241 TTTTATTATTAATTAATTTCAAGTTTCCAGATATTAGTATATAATTTTAAAGAAATTTGAA 300  
 DB 241 TTTTATTATTAATTAATTTCAAGTTTCCAGATATTAGTATATAATTTTAAAGAAATTTGAA 300

QY 301 CAACACCAAGTTGTTTAAACAGGTTGTTAAATCAGAGATTTACTTTAAAGGAAAGATGTT 360  
 DB 301 CAATATCAATTTGTTTAAACAGGTTGTTAAATCAGAGATTTACTTTAAAGGAAAGATGTT 360

QY 361 GACCAATATCCTCTCTCAAGAGATATCAACAGAAATCTTTGATTTTAAAGCAAAA 420  
 DB 361 GAAACAATATCCTCTCTCAAGAGATATCAACAGAAATCTTTGATTTTAAAGCAAAA 420

QY 421 TTTTGAAGTCTATTTGCTGAAACAGCTTTTCCAGGCTTTTCAAGAAAGTCGTCCT 480  
 DB 421 TTTTGAAGTCTATTTGCTGAAACAGCTTTTCCAGGCTTTTCAAGAAAGTCGTCCT 480

QY 481 ATTTTAAACAGGAGTTTCATATTGTTATTAGTAATCATATAAGATTTTAAACGAGTAGCGACT 540  
 DB 481 ATTTTAAACAGGAGTTTCATATTGTTATTAGTAATCATATAAGATTTTAAACGAGTAGCGACT 540

QY 541 GACTCTCATCGTATGAGCCAAAGCTTTAATCACTTTTGGCAATATCTTTCAGCAGATTTGATG 600  
 DB 541 GACTCTCATCGTATGAGCCAAAGCTTTAATCACTTTTGGCAATATCTTTCAGCAGATTTGATG 600

QY 601 GTAGTTCTTCCAGTAAATCTTTGAGAGAAATTTTTCAGCAGATTTTTCAGATGATATTGAG 660  
 DB 601 GTAGTTCTTCCAGTAAATCTTTGAGAGAAATTTTTCAGCAGATTTTTCAGATGATATTGAG 660

QY 661 ACCGTTGAGTATTTTCTTCCAGGCAAAATCTTTCAGCAGATTTTTCAGATGATATTGAG 720  
 DB 661 ACTGTAGAGATTTTCTTTCAGGCAAAATCTTTCAGCAGATTTTTCAGATGATATTGAG 720

QY 721 TATACACGCTCTTTAGAGGAAATTTATCCCATATACAGACCTTTTATTAATGACAGAAATTT 780  
 DB 721 TATACGCTCTTCTTAGAGGAAATTTATCCCATATACAGACCTTTTATTAATGACAGAAATTT 780

QY 781 GAGACGAGGTTGTTTTCATACCAATCCCTTCCGACGCTATGGAACGTCCTTCTTCTT 840  
 DB 781 AACACTACTATTACTTTTAAATGCTGTAACCTTACGCCAGTCAATGAGCGTCCGCTCTT 840

QY 841 ATTTCTAATGCTACTCAAAATGGTACTTAAAGCTTTGAGATTTACTCAAAATCATATTCTCA 900  
 DB 841 TTATCAAGTGGATCAAAATGGTACTTAAAGCTTTGAGATTTACTCAAAATCATATTCTCA 900

QY 901 GCTCATGTTAACTCACCTGAGGTTGGTAAGGTAAACGAGGATTTAGATATTGTTAGTTCAG 960  
 DB 901 GCCCATGTTCACTCTCCAGAGTTGGTAAAGTAAACGAGGATTTAGATATTGTTAGTTCAG 960

QY 961 TCTGCTAGTATTTAACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 1020  
 DB 961 ACTGCTGAGATTTGACCAATTTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 1020

QY 1021 ATTTAAAGTGAACAGTAAATAATTTCTTTCTTATCACCAGTTCCAGCAATTCACCTTAACA 1080  
 DB 1021 TTTAATAGCGAAAGGTGACCAATTTAGCTTTTATCTCAGCTGTTCTGCTTCTTCTTCTG 1080

QY 1081 CCAGGCAATGAGGAGAAAGTTTATCCAAATTAATTTACAGCAGTACGAGCAAA 1133  
 DB 1081 CCAGCAGATACTGCAAGAGCTTCAATGAGCTCATTAACAGGTTCTGTTCTTCTTCTG 1133

RESULT 15

ABA90521\_00

WP	Sequence	split	into	24	fragments	LOCUS	ABA90521	Accession	ABA90521
WP	Fragment	Name	Begin	End					
WP	ABA90521_00		1	110000					
WP	ABA90521_01		100001	210000					
WP	ABA90521_02		200001	310000					
WP	ABA90521_03		300001	410000					
WP	ABA90521_04		400001	510000					
WP	ABA90521_05		500001	610000					
WP	ABA90521_06		600001	710000					
WP	ABA90521_07		700001	810000					
WP	ABA90521_08		800001	910000					
WP	ABA90521_09		900001	1010000					
WP	ABA90521_10		100001	1110000					
WP	ABA90521_11		1100001	1210000					
WP	ABA90521_12		1200001	1310000					
WP	ABA90521_13		1300001	1410000					
WP	ABA90521_14		1400001	1510000					
WP	ABA90521_15		1500001	1610000					
WP	ABA90521_16		1600001	1710000					
WP	ABA90521_17		1700001	1810000					
WP	ABA90521_18		1800001	1910000					
WP	ABA90521_19		1900001	2010000					
WP	ABA90521_20		2000001	2110000					
WP	ABA90521_21		2100001	2210000					
WP	ABA90521_22		2200001	2310000					
WP	ABA90521_23		2300001	2365589					



XX PF 21-MAR-2001; 2001WO-US009180.  
XX PR 21-MAR-2000; 2000US-0191078P.  
XX PR 23-MAY-2000; 2000US-0206848P.  
XX PR 26-MAY-2000; 2000US-0207727P.  
XX PR 23-OCT-2000; 2000US-0242578P.  
XX PR 27-NOV-2000; 2000US-0253625P.  
XX PR 22-DEC-2000; 2000US-0257931P.  
XX PR 16-FEB-2001; 2001US-0269308P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Haselbeck R, Ohlsen KU, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
XX PI Yamamoto RT, Xu HH;  
XX DR WPI; 2001-611495/70.  
XX DR P-PSDB; AAU35314.  
XX PT New polynucleotides for the identification and development of  
XX PT antibiotics, comprise sequences of antisense nucleic acids.  
XX PS Claim 27; SEQ ID NO 6810; 511pp; English.  
XX CC The invention relates to antisense inhibitors of genes essential to  
XX CC prokaryotic cellular proliferation, their use in identifying the genes,  
XX CC their use in the discovery of novel antibiotics, the essential genes  
XX CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
XX CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
XX CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
XX CC useful for the identification of potential new targets for antibiotic  
XX CC development. The antisense nucleic acids can also be used to identify  
XX CC proteins used in proliferation, to express these proteins, and to obtain  
XX CC antibodies capable of binding to the expressed proteins. The proteins can  
XX CC be used to screen compounds in rational drug discovery programmes. The  
XX CC antisense nucleic acid sequence is also useful to screen for homologous  
XX CC nucleic acids which are required for cell proliferation in a wide variety  
XX CC of organisms. The present sequence encodes an essential prokaryotic  
XX CC cellular proliferation protein. Note: The sequence data for this patent  
XX CC did not form part of the printed specification, but was obtained in  
XX CC electronic format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 1131 BP; 385 A; 198 C; 193 G; 355 T; 0 U; 0 Other;  
Query Match 30.1%; Score 341.2; DB 4; Length 1131;  
Best Local Similarity 57.0%; Pred. No. 3.2e-63;  
Matches 644; Conservative 0; Mismatches 483; Indels 3; Gaps 1;  
QY 4 ATTCAATTTTCAATTAATGCGACATTTATTTATGCTTTTAAATACAACTAAACGCTCT 63  
Db 1 ATGAAATTAACGGTCAACGAAGTCTTTTACGAAGATTAACAATGTTCAACGAGCA 60  
QY 64 ATTACACTAAATGCGATTCCTTATCTTCAATCAATTAATTAATTAATTAATTAATTA 123  
Db 61 ATTCTCTTAAACCAACGATTCCTTATCTTCAATCAATTAATTAATTAATTAATTAAT 120  
QY 124 GGAGTAATCTTAAACAGGCTCAACGGTCAATATCAATTAATTAATTAATTAATTAATTA 183  
Db 121 GGCATTATCACTACTGAGGAGTAACCGGATTAATTTCAATTAATTAATTAATTAATTA 180  
QY 184 AATGAATGCTGGTTTCTTAATACCTCTCCAGGAGCTATTTTATTAAGAGCTAGTTT 243  
Db 181 GATGAAATGCTCAATGACCAATGACGACGATGATGATGATGATGATGATGATGATGAT 240  
QY 244 TTTATTAATATTTTCAAGTTTCCAGATATTAGTATAATGTTAAAGAAATTAAGAA 303  
Db 241 TTTGCGGAATATTCGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 300  
QY 304 CACCAAGTTGTTTAAACAGGTTGTTAAATTCAGAGATTACCTTAAAGGAAAGATGTTGAC 363  
Db 301 AATCAAGTTGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 360

QY 364 CAGTATCTCTGCTTCAAGAGAGTATCAACAGAAATTCCTTTGATTTTAAACCAAAATTA 423  
Db 361 AATTATCCGCACTTACCAGTAATTCATCTCAAAACCAATTAATTAATTAATTAATTAAT 420  
QY 424 TTGAAGTCTTATTTTCTGTAACAGCTTTTGCAGCAGGTTTCAAGAAAGTCTCTATT 483  
Db 421 TTGACAAAATTTATTTAGTGAACAGGTTTCTGTATCGATGACGAAAGTCTGCAAT 480  
QY 484 TTAACAGGAGTTTCAATTTGATTAATTAAGATTTTAAAGCAGTACGAGCTGAC 543  
Db 481 TTAACCTGGGTTTCACTTTTATTTAGAAATCAAAA---TTACTTGGCGTTGCGACAGAT 537  
QY 544 TCTATCTGATGAGCCCAAGTTTAACTACTTTTGGCAATATCTTCAAGAGATTTGATGTA 603  
Db 538 TCACATCTGTTTAAAGTCAACGTTGATCCGACAGAAACAAAGCAGTAGAAGACTTTAAACATT 597  
QY 604 GTTCTTCCAAAGTAATCTTTGAGAGAAATTTTTCAGCAGTATTTTACAGATGATTTGAGACC 663  
Db 598 GTAATTCAGGAAAGTTTAACTGAACTTTCTCGTTCAITTAACCAATGAAGAAAGAAATG 657  
QY 664 GTTGAAGTATTTTCTTCAACCAAGCCAAATCTTGTTCAGAAAGTGAACACATTTCTTTTAT 723  
Db 658 GTTGAATTCAGCATTTATGGAACCAAGTCTATTTTAAACAGAGAAACAAATGCTCTCTAT 717  
QY 724 ACAGCTCTTTAGAGGAAATTTATCCGATACAGACCGTTTATTAATTAATGAAGAAATTTGAG 783  
Db 718 TCTGTTTGTGTAAGGAAATTTATCTGATACCAACCGTTTAAATTTCCAACTAGCCATAAC 777  
QY 784 ACGAGGTTGTTTCAATCCCAATCCCTTTCGACCGCTATGGAACGTCCTTCTTCTGATT 843  
Db 778 ACGCAATTTGAATTTTATGTACAGAAATGCTTTTCAAGATTCGAACTGCTCTTACTT 837  
QY 844 TCTAATCTACTCAAAATGCTACTTTAAGCTTGAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 903  
Db 838 TCTATGAAGAGAGTAAACAAATTTGTTCGCTTTTCAATTTTCAACAGATCTCTGTGTTTA 897  
QY 904 CATGTTAACTCACCTGAGGTTGTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 963  
Db 898 TATGGAATTTCACTGAAATTTGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 957  
QY 964 GGTAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1023  
Db 958 GGTGAAGCGTTGATATTTCTTCAACCCAGATTTATATGAAGATGCGTTGCGAGCGTTT 1017  
QY 1024 AAAAGTGAACAGTAAATTTCAATTTCTTATCAGGATTCGACCATTCACCCCTAACACCA 1083  
Db 1018 GCGATATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1077  
QY 1084 GCGATGAGGAAAGAAATTTTATCCAAATTAATTAATTAATTAATTAATTAATTAATTAAT 1133  
Db 1078 ACCGAAACAGAACTAGATTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1127

RESULT 17  
AC32930  
ID AC32930 standard; DNA; 1128 BP.  
XX AC32930;  
AC AC32930;  
XX AC32930;  
DT 19-JUN-2003 (first entry)  
XX Prokaryotic essential gene #14587.  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
XX drug design; gene..  
XX Enterococcus faecalis.  
XX WO20027183-A2.  
XX 03-OCT-2002.  
XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
DR P-PSDB; ABU29060.  
XX New antineoplastic nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 14; SEQ ID NO 20800; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1128 BP; 383 A; 196 C; 193 G; 356 T; 0 U; 0 Other;  
Query Match 29.8%; Score 338; DB 8; Length 1128;  
Best Local Similarity 56.8%; Pred. No. 1.5e-62;  
Matches 642; Conservative 0; Mismatches 485; Indels 3; Gaps 1;  
QY 4 ATTCAATTTTCAATTAATCGCATATTTATTTCATGCTTTTAAATACAACTAAACGTCT 63  
DB 1 ATGAAATTAACGGTCAAGGAAAGTGTCTTTTACAAAGAAATTAACAACCTGTTCAACGAGCA 60  
QY 64 ATTAGCACTAAATAATGCCATCTCTTCTTCAATCAATAAAATTTGAAGTCACTTCTACA 123  
DB 61 ATTCTTCTTAAACAACGATTCATCTTAACAGGTGTATAATTTGTGTATCAGAAGAT 120  
QY 124 GGAGTAACCTTTAAACGGGTCTAACGGTCAAAATATCAATTTGAAACAACTATTCCTCTAAGT 183  
DB 121 GGCTTATCACTTACTCTGGGAGTAACGGGATATTTCAATTTGAAAGTTTTTAAAGTAAAGAT 180  
QY 184 AATGAAATGCTGGTTTGGTAAATACCTCTCCAGGAGCTATTTTATTAGAAGCTAGTTT 243

Db 181 GATGAAAAGCTCAAAATGACCAATGAAACGCACAGGTAGCATCGTTTTTACAATCTCGTTTC 240  
QY 244 TTTATTAATATTTTCAAGTTTGGCAGATATTTAGTATATAAATGTTTAAAGAAATTTGAACAA 303  
Db 241 TTTGCGGAAATTTTCGTAATATACCAAGAGATATGTTTCAAAATGGAAGTTCTAGATAAC 300  
QY 304 CACCAAGTTGTTTTAAACCAAGTGGTAAATCAGAGATTTACCTTTAAAGGAAAGATGTTGAC 363  
Db 301 AATCAAGTTGCAATTTACTTTCAGGAAAGCTGATTTTACGGTTAAATGGAATAGATCTGAT 360  
QY 364 CAGTATCTCGTCTACAAAGAGATATCAACAGAAAAATCTTTGATTTTAAACCAAAATTTA 423  
Db 361 AATATCCGCACTTACCAAGTAAATGATACTCAAAACCAAAATGAAATTTACCTGTTCAATTA 420  
QY 424 TTGAAGTCTATTTATGCTGAAACAGCTTTTGCAGCCAGTTTTCACAAAGAGTCTGCTATTT 483  
Db 421 TTGCAAAAATTTATTTAGTGAACAGGTTTTGCTGATCGATGCACGAAAGTCTGCAATTT 480  
QY 484 TTAACAGAGTTTCATATTTGATTAAGTAAATCATTAAGATTTTAAAGCAGTAGCGACTGAC 543  
Db 481 TTAACGGGTTCTACTTTATTTTAAAGAAATCAAAAA---TTACTTTGCCGTTGCGCAGAT 537  
QY 544 TCTCATCTGATGAGCAACGTTTAAATCACTTTGGACAATCTTTCAGCAGATTTGATGGTA 603  
Db 538 TCACATCGTTTAAAGTCAACGTTGATCCGACAGAACAGCAGTAGAAGACTTTAAACATTT 597  
QY 604 GTTCTTCAAGTAAATCTTTGAGAGAAATTTTTCAGCAGTATTTTACAGATGATATTTGAGACC 663  
Db 598 GTAATTCAGGAAAGGTTTAACTTCTGTTCAATTAACCAATGAAGAAGAAATG 657  
QY 664 GTTAGGATTTTCTTCAACAGCCAAATCTTTGTCAGAGTGAACACATTTCTTTTAT 723  
Db 658 GTTGAATCAGCATTTATGGAATCAAGTCTATTTTAAACAGAGAAACAAATGTAATTTCTAT 717  
QY 724 ACAGCCTCTTTAGAGGAAATTTATCCCATACAGACCGTTTATTAATGACAGAAATTTGAG 783  
Db 718 TCTGTTTGTAGAGGAAATTTATCCTGATACCAACCGTTTAAATTCACACTAGCCATTAAC 777  
QY 784 ACAGGAGTTGTTTCAATACCAATCCCTTCCGCCACGCTATGGAAACGTCCTTCTTGATTT 843  
Db 778 ACGCAATTTGAATTTTATGTACCAAGATTTGCTTTTTCAGCAATCGAACGTCCTTCTTACTT 837  
QY 844 TCTAATGCTACTCAAAATGGTACTGTTTAAAGCTTGAAGTACTTCAAAATCATATTTTCAGCT 903  
Db 838 TCTCATGAAGACGTAACAATTTGTCGCTTTTCAATTTTCCAGAGATTTCTGTTGTTTA 897  
QY 904 CATGTTAACTCACCTGAGGTTTGGTAAGTAAACAGGAGTTTATAGATATTTAGTCACTCT 963  
Db 898 TATGAAATTTCACTTGAATTTGGAAGAGTTCGAGAGCTTTTAACTATGAAATGTTTCT 957  
QY 964 GGTAGTGAATTTAACTATCAGCTTCAATCCAACTTACCTTATTTAGTCTTTTAAAGCTATT 1023  
Db 958 GGTGAAGCTTGGATATTTCTTTCAACCCAGATATATGAAAGATGCGTTGCGAGCGTTT 1017  
QY 1024 AAAAGTGAACAGTAAATAATTTCTTATCAGCAGTTTCAGCCATTTACCCCTAACACCA 1083  
Db 1018 GCGGATATGAATATTTACCGTGAATTTCTTTTCCAAATTCGTTCCGTTTACATTTGAGGCCA 1077  
QY 1084 GCGGATGAGGAAAGGTTTATCCAAATTAATTTACACAGTACGAAACAAA 1133  
Db 1078 ACCGAAACAGAACTAGATTTTCACTCACTAATTTACACCGGTTGCTACAAA 1127

RESULT 18  
AA13559  
ID AA13559 standard; DNA; 8001 BP.  
XX  
AC AA13559;  
XX  
DT 19-MAR-1999 (first entry)  
XX  
DE Enterococcus faecalis genome contig SEQ ID NO:622.  
XX



```

PF 04-MAY-1998; 98US-00070927.
XX
PR 04-MAY-1998; 98US-00070927.
XX
PA (KUNSCH C A.
PA (DILLON P J.
PA (BARASH S.
XX
PI Kunsch CA, Dillon PJ, Barash S;
XX WPI; 2002-750065/81.
XX
PT Computer readable medium having recorded on it a Enterococcus faecalis
PT nucleotide sequence useful for detecting diseases related to Enterococcus
PT infections in animals.
XX
PS Claim 1; Page; 119pp; English.
XX
CC The present invention relates to a new computer readable medium with an
CC Enterococcus faecalis nucleotide sequence. The invention is useful to
CC diagnose the presence of E. faecalis in a sample or determining the
CC presence of a specific microbe in a sample. The invention is also useful
CC for modulating the growth or pathogenicity of E. faecalis, in a vaccine to
CC confer resistance to Enterococcal infection, for commercial, therapeutic
CC and industrial purposes, and for fermenting a particular sugar source or
CC to produce a particular metabolite. The invention is useful for detecting
CC diseases related to Enterococcus infections in animals, and for detecting
CC E. faecalis using biotech technology. The present nucleic acid sequence
CC represents an Enterococcus faecalis contig DNA sequence of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from USPTO
CC at http://seqdata.uspto.gov
XX
SQ Sequence 8001 BP; 2672 A; 1407 C; 1670 G; 2250 T; 0 U; 2 Other;

Query Match 29.2%; Score 330.6; DB 6; Length 8001;
Best Local Similarity 56.9%; Pred. No. 8.1e-61;
Matches 645; Conservative 0; Mismatches 485; Indels 4; Gaps 2;

QY 1 ATGATTTCAATTTTCAATTAATCGCACATTTATTTATTTTCATGCTTTTAAATACAACTAAACGTT 60
DB 598 ATTATGAAATTAACGGTCNACGAAGTGTCTTTTACAGAAATTAACAACTGTTTCAACGA 657

QY 61 GCTATTAGCACTAAATAATGCCATTCCTATTCCTTT-CATCAATAAATAATGAAGTCACATTC 119
DB 658 GCAATTTCTTCTTAAACAAACGATCCCAATCTTAACAGGTGTAAATAATGTGTTATCAGA 717

QY 120 TACAGGAGTAACTTTTAACAGGGTCTTAACGGTCAATATATCAATTAATGAACACTATTCCTGT 179
DB 718 AGATGGCTTATCACTTACTGGGAGTAAACGGGATATTTCAATTTGAAAGTTTTTTAAGTAA 777

QY 180 AAGTAATGAAATGCTGGTTTCTGCTAATTAACCTCTCCAGGAGCTATTTTATTAGAAGCTAG 239
DB 778 AGATGATGAAAGGCTCAATGACCATTAACGACAGGAGTACGATCGTTTACATCTCG 837

QY 240 TTTTATTAATATTTATTTCAAGTTTGGCAGATATTAATGATAAATGTTAAAGAAATGA 299
DB 838 TTTCTTTGGCGAAATTTATTCGTAATTTACAGAGATATGTTCACAATGGAAGTTCTAGA 897

QY 300 ACAACACCAAGTTGTTTAAACAGTGGTAAATACAGATTAACCTTAAAGGAAAGATGT 359
DB 898 TAAACATCAAGTTGCAATTTACTTACGGAAAGGCTGATTTTACGGTTAATGGAATGATGC 957

QY 360 TGACCATGATTCCTGCTACAGAGAGTATCAACAGAAATCCCTTTGATTTTAAACAAA 419
DB 958 TGATAATATCGCACTTACCAAGTAATGATATCAACCAACAAATGAATTAACCTGTCA 1017

QY 420 ATTATTAAGTCTATTATTTGCTGAACAGCTTTTGCAGCCAGTTTACAGAAAGTCGTC 479
DB 1018 TTTATTTGACAAATTTATTTAGTGAACAGGTTTGTCTGATCGATCGACGAAAGTCGTC 1077

QY 480 TATTTTACAGAGTTCATTTGTTATTAAGTAATCATTAAGATTTTAAAGCAGTAGCGAC 539
DB

```

```

DB 1078 AATTTTAACTGGGGTTCACTTTATTTTAGAAAAATCAAAAA---TTACTTGGCGTTGGAC 1134
QY 540 TGACTCTCATCGTATAGCCCAACGGTTTAAATCACTTTTGGACAATACTTCAGCAGATTTGAT 599
DB 1135 AGATTTCAATCGTTTAAAGTCAACGTGTGATCCCGACAGAACAGCAGTAGAAGACTTTAA 1194
QY 600 GGTAGTTCTTCCAAGTAAATCTTTGAGAGAAATTTTTCAGCAGTATTTTACAGATGATATGA 659
DB 1195 CATTTGTAATTCAGGAAAAAGTTTAACTGAACTTTTCTCGTTTCATTTAAACCAATGAAGA 1254
QY 660 GACGGTTCAGGTATTTTCTCACCAAGCAAAATCTTTTTCAGAACTGGAACACATTTCTTT 719
DB 1255 AATGTTGAAATCAGCATTTTGAATAATCAAGTGTATTTTAAACAGAAACAACTGACTT 1314
QY 720 TTATACAGCCCTCTTTAGAAAGAAATTTATCCGATAACAGACCGTTTATTAATGACAGAAAT 779
DB 1315 CTATTTCTGTTTGTAGAAAGAAATTTATCTGTATACCAACCGTTTAAATTTCCAACTAGCCA 1374
QY 780 TGACAGCGAGGTGTTTTCATATACCAATCCCTTCGCCACGCTATGGAAACGTCCTTCTT 839
DB 1375 TAAACAGCAAAATTTGAATTTTATGTACCAAGAAATTTGCTTTTTCAGCAATCGAACGTCCTT 1434
QY 840 GATTTCTAATGCTACTCAAAATGGTACTGTTTAAAGCTTGAGATTACTCAAAATCATATTTTC 899
DB 1435 ACTTTCTCATGAGGACGTAACATATTTGTCGCTTTCAATTTTCCAGATTTCTGTTGT 1494
QY 900 AGCTCATGTTAACTCACTGAGGTGTTGTAAGTAAACGAGGATTTAGATATTTGTTAGTCA 959
DB 1495 TTTATATGAAATTTTACCTGAAATTTGGAAGAGTTCGAAAGCTTTTAAACTATGAAATGT 1554
QY 960 GTCTGGTAGTGAATTAACATATCAGCTTCAATCCAACTTACCTTATTTAGTCTTTTAAAGC 1019
DB 1555 TTCTGGTGAAGCGTTGGATATTTCTTCAACCCAGATTTATGAAAGATGCGTTCGGAGC 1614
QY 1020 TATTAAGAGTGAACAGTAAATAATTTCAATTTATCACCAGTTTCAGCAATTCACCCCTAAC 1079
DB 1615 GTTTGGCATATGAATATTTACCGTGAATTTCTTCTCAATTCGTCGTTTACATTGGA 1674
QY 1080 ACCAGGCGATGAGGAGAAAGTTTATCCAAATTAATTAACACAGTACGAAACAAA 1133
DB 1675 GCCAACGAAACAGAACTAGATTTTCAATTAATTAACCGGTTTCGTACAAA 1728

RESULT 20
ADC91275
ID ADC91275 standard; DNA; 1155 BP.
XX
AC ADC91275;
XX
DT 01-JAN-2004 (first entry)
XX
DE E. faecium DNA sequence SEQ ID 902.
XX
KW ds; gene; urinary tract infection; bacteraemia; endocarditis; wound;
KW abdominal-pelvic infection.
XX
OS Enterococcus faecium.
XX
PN US6583275-B1.
XX
PD 24-JUN-2003.
XX
PF 30-JUN-1998; 98US-00107532.
XX
PR 02-JUL-1997; 97US-0051571P.
PR 14-MAY-1998; 98US-0085598P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI; 2003-799836/75.
DR P-PSDB; ADC94929.

```







for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

Claim 14: SEO ID NO 21823: 1766pp: English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: the sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published/pct](http://wipo.int/pub/published/pct) sequences

Sequence 1128 BP: 355 A: 210 C: 212 G: 351 T: 0 U: 0 Other: 0

Query Match	27.8%	Score	315.6	DB	8	Length	1128
Best Local Similarity	55.6%	Pred.	No. 9.2e-58				
Matches	628	Conservative	0	Mismatches	499	Indels	3
Gaps	1						
Qy	4	ATTCAATTTTCAATTAATCGCACAATTATTTATTTATTCATGCTTTAAATACAACTTAACAGTGC	63				
Db	1	ATGAAGTTACTTTAAACCGGAGCTAGCTTTATGCAAGAAATTGCAAACTGTTTCAACGAGCT	60				
Qy	64	ATTAGCACTAAAAATGCAATTCCTTATTTCTTTTCATCAATAAAAAATTGAAGTCACCTTCTACA	123				
Db	61	ATTTCAGCAAAACCAAGATCCCTTATTTTGACAGGTGTAAAAATCACACTGACACAAGAA	120				
Qy	124	GGAGTAATCTTTAACAGGGTCTAACGGTCAAAATATCAATTTGAAACACATATTCCTGTGAAGT	183				
Db	121	GGTTTGACTTTGACGGGAGCAACGCTGATATATCAATTGAAACTTTTTTGTCTGTGTGAA	180				
Qy	184	AATGAAATGCTGCTTTGCTTAATTTACCTCTCCAGGAGCTATTTTATTAGAGCTAGTTTTT	243				
Db	181	AACGAAAGCAAAATATGCAAAATCGAATCTACTGGTCCATGTGTTTTTACAAGACAGTTTC	240				
Qy	244	TTTATTAATATTTATTTCAAAGTTTCCAGATATTAGTATAAATGTTTAAAGAAATTTGAACAA	303				
Db	241	TTTAGCGAAATCAITTCGGAGACTTTCCTGAGAAACATTTACTTTAGAAGTTTTAGAAAT	300				
Qy	304	CACCAAGTTGTTTTAACACAGTGGTAAATCAGAGATTACCTTAAAGGAAAGATGTTGAC	363				
Db	301	AAACAAGTAGCGATCACTTCTGGAAGAGCGAAATTTTATCGTAAATGGATTAGTCAGAT	360				
Qy	364	CAGTATCCTCGTCTACAAGAGTATCAACAGAAATCCTTTGATTTTAAAAACAAATTA	423				
Db	361	AACATATCTCATCTTCTGTTGCGAAGGCCATACCCAGATGAATTTACCTGTACAGTA	420				
Qy	424	TTGAAGTCTATTTATTTGCTGAAACAGCTTTTTTGACGCCAGTTTTACAAGAAAGTCGTCCTATT	483				

XX PA (ELIT-) ELITRA PHARM INC.  
 XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 XX PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;  
 XX PI WPI; 2003-029926/02.  
 DR P-PSDB; ABU42909.  
 XX PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX PT  
 PS Claim 14: SEQ ID NO 34649; 1766pp; English.  
 XX PT  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid;  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway; (8)  
 CC required for proliferation, or that inhibits cellular proliferation; (9)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ  
 Sequence 1131 BP; 421 A; 175 C; 186 G; 349 T; 0 U; 0 Other;  
 Query Match 21.8%; Score 247.4; DB 8; Length 1131;  
 Best Local Similarity 52.9%; Pred. No. 3.3e-43;  
 Matches 580; Conservative 0; Mismatches 511; Indels 6; Gaps 2;  
 QY 1 ATGATTCATTTTCAATTAATCGCACATTATTATTTCATGCTTTTAATACAACTAAACGT 60  
 Db 1 ATGATGGAATTCACAAATTAAGAGATTTATTTTAAATCACTTAACGACACATAAAG 60  
 QY 61 GCTATTAGCACTAAATATGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 120  
 Db 61 GCTATCTCACCTAGAACAACTTTACCGATTTTAAACAGGATCAAAATTTGATGCTTAAGAA 120  
 QY 121 ACAGAGATTAATTAACAGGCTTAACGGTCAATATCAATTAATTAATTAATTAATTAATTAAT 177  
 Db 121 AATGAGTCATTTTAACTGGGTGAGATTCAGAAATATCAATTAATTAATTAATTAATTAATTAAT 180  
 QY 178 GTAAGTAATGAAATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 237  
 Db 181 CAAGTTGATGCTGAAGAAATGTTGAATTAACAGAACAGATCAGTTGATGCTGGTGGTGGTGGTGGT 240  
 QY 238 AGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 297  
 Db 241 CGTTTCTCGTTGATATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 300

QY 298 GAACAAACCAAGTTGTTTAAACAGTGGTAAATCAGAGATTACCTTAAAGGAAAGAT 357  
 Db 301 GAACAAATCCAAACGTTAATCATCAGGTCAATCAGAAATTAACTTAAGTGGCTTAGAT 360  
 QY 358 GTTGACCAAGTATCCTCGTCTACAGAAATATCAACAGAAATCCCTTCAATTTTAAACAA 417  
 Db 361 CCTGATCAATACCCATTTATTTACCTGAGGTATCAGAGATGACGCTTATCAATTTGCTGTT 420  
 QY 418 AAATTTTGAAGTCTTATTTATTTGCTGAAACAGCTTTTGCAGCCAGTTTACAAAGAAAGTCT 477  
 Db 421 AAGGTACTTAAATAATATCATTTGCAAACTAATTTTGCAGTGTCCACCTCAGAAACACGA 480  
 QY 478 CCTATTTTAAACAGAGGATTCATATTTGTTATTAAGTAAATCAATAAGATTTTAAAGCAGTAC 537  
 Db 481 CCAGTACTTACTGGTGTAACTGGCTTATA---CAAGATAATGAATTAATATGCACAGCA 537  
 QY 538 ACTGACTCTCATCGTATGAGCAACGTTTAAATCACTTTGGACAATCTTTTCAGCAGATTG 597  
 Db 538 ACAGATTACACCGCTTAGCTGTAAAGAAAGTTACAGTTAGAAATGAATCAGAAATAA 597  
 QY 598 ATGTTAGTCTTCCAAAGTAAATCTTTGAGAGAAATTTTCAGCAGTATTTTACAGATGATATT 657  
 Db 598 AATGTCATCATCTCTGTTAAAGCTTTTCTGAATTAACAAATTTATGAGTGACAGCGAC 657  
 QY 658 GAGACCGTTGAGGTATTTTCTCACCAGCCAAATCTTTGTCAGAAAGTGAACATTTCT 717  
 Db 658 GAAGATATTGATATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 717  
 QY 718 TTTTATACACGCTCTTTAGAGGAAATTTATCCCGATACAGACCGTTTATTAATGACAGAA 777  
 Db 718 TTATCTCAGCTTTTACTTTGAAGGTCATTTCCAGATACGACAGCTTTTATTTCCCGAGAAAT 777  
 QY 778 TTTGAGAGGAGGTTGTTTTCATATACCAATTCCTTCCGCAACGCTATGGAACGTCCTTC 837  
 Db 778 TATGAGATTAAATTTAGGAATTTAAATGAGGACTTCTATCATGCAATTTGATCGTGCATCT 837  
 QY 838 TTGATTTCTAATGCTACTCAAAATGTTGTTTAAAGCTTTGAGATTACTCAAAATCATTT 897  
 Db 838 TTATTACAGTGAAGTGGAAATTAATGTTTAAATTAAGTACAGGTAATGAATTAAGTT 897  
 QY 898 TCAGTCTCATGTTAACTCACCTGAGTTGGTAAAGGTAACGAGGATTTAGATATTTGTTAGT 957  
 Db 898 GAACTTTTCATCTACTCTCTGAAATTTGGTACTGTTTAAAGAAAGTTAACGCTAATGAT 957  
 QY 958 CAGTCTGTTAGTATTAATCACTATCAGCTTCAATCCAACTTACCTTATGAGTCTTTAA 1017  
 Db 958 GTAGAGGCGGAAACTTTGAAATTTCTTCACTCAAAATACATGATGATGCTTTTAA 1017  
 QY 1018 GCTATTTAAAGTGAACAGTAAATAATTTCTTCTTATCACCAGTTCGACCATTTACCCCTA 1077  
 Db 1018 GCCATTGATTAATGATGAGTAGAGTAGAATTTCTTTGGTACATGAACCATTTATCTTA 1077  
 QY 1078 ACACGAGCGCATGAGGA 1094  
 Db 1078 AAACCAAAAGATGATGA 1094

RESULT 23  
 ABN92866  
 ID ABN92866 standard; DNA; 1158 BP.  
 XX ABN92866;  
 AC AC  
 XX XX  
 XX XX  
 DT 24-JUL-2002 (first entry)  
 DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:2329.  
 XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 KW antibacterial; gene therapy; gene; ds.  
 XX Staphylococcus epidermidis.  
 XX US6380370-B1.  
 PN

XX 30-APR-2002.  
 XX PF 13-AUG-1998; 98US-00134001.  
 XX PR 14-AUG-1997; 97US-0055779P.  
 XX PR 08-NOV-1997; 97US-0064964P.  
 XX PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX PI Doucette-Stamm LA, Bush D;  
 XX DR WPI; 2002-381255/41.  
 XX DR P-PSDB; ABP40321.  
 XX PT Novel isolated nucleic acid encoding a Staphylococcus epidermis  
 XX PT polypeptide, useful for diagnosing and treating bacterial infections.  
 XX PS Disclosure; SEQ ID NO 2329; 267pp; English.  
 XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences can  
 CC also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life cycle  
 CC or inhibit S. epidermidis infection. N.B. The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from the USPTO web site  
 XX  
 SQ Sequence 1158 BP; 432 A; 177 C; 192 G; 357 T; 0 U; 0 Other;  
 Query Match 21.8%; Score 247.4; DB 6; Length 1158;  
 Best Local Similarity 52.9%; Pred. No. 3.3e-43;  
 Matches 580; Conservative 0; Mismatches 511; Indels 6; Gaps 2;  
 QY 1 ATGATTCAATTTTCAATTAATCGACATTAATTTATTTATTCATGCTTTTAAATACAACTAAACGT 60  
 DB 25 ATGATGGAATTCACAAATTAAGAGAGATTTATTTATTAATCAACTTTAAACGACATTAAG 84  
 QY 61 GCTATTAGCATAAATGCCATTCCTATTCCTCATCAATTAATAATGAGTCACTTCT 120  
 DB 85 GCTATCTCACCTAGAACAACTTTACCGATTTTAAACAGATGATCAAAATTTGATGCTAAAGAA 144  
 QY 121 ACAGAGATTAACCTTTAAACAGGGTCTAACCGTCAATATCAATTTGAAACACATTTCC--T 177  
 DB 145 AATGAAGTCATTTTAACTGGGTGAGATTCAGAAATATCAATAGAAATCACTATTCCAAA 204  
 QY 178 GTAAGTAATGAATGCTGGTTGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 237  
 DB 205 CAAGTTGATGGTGAAGAAATTTGAAATTTACAGAAACAGGATCAGTTGCTTCTGCTGT 264  
 QY 238 AGTTTATTTATTAATTAATTTTCAAGTTTGGCAGATATAGTATTAATTAATTAATTAATTAATTA 297  
 DB 265 CGTTTCTGTTGATATTAATTAATTAATTAATTTCTCGAAAGAGATTTAAATTTATCACTAAT 324  
 QY 298 GAACAAACACCAAGTTGTTTAAACAGGTGTAATCAGAGATTACCTTTAAAGGAAAGAT 357  
 DB 325 GAACAAATTCCAACGTTAATCAATCAGTCATTCAGATTTTAACTTAAGTGGCTTAGT 384  
 QY 358 GTTGACCAAGTATCCTGGCTTACAAGAGATPATCAACAGAAAAATCCTTTGATTTTAAAGACA 417  
 DB 385 CCTGATCAATACCCATTTATTAATCAGGATGATCAGGATGATGATGATGATGATGATGATGATGAT 444  
 QY 418 AATATTATTAAGTCTATTTATTTGCTGAAACAGCTTTTGGACGAGTTTACAGAAAGTGT 477  
 DB 445 AAGGTACTAAAAAATATATTCGACAAACTTAATTTTTCAGTGTCCACCTTCAGAAACACGA 504  
 QY 478 CCTATTTTAAACAGGAGTTCATATTTGATTAAGTAATCATATAAGATTTTAAAGCAGTAGGG 537  
 DB 505 CCAGTACTTACTGTTGGTAAACTGGCTTATA---CAAGATTAATGAATTAATATGCACAGCA 561

QY 538 ACTGACTCTCATCGTATGAGCCAAAGTTTAAATCACTTTTGGACAATACTTTGACGAGATTTG 597  
 DB 562 ACAGATTACACACCGCTTAGCTGTAAAGAAAGTTACAGTTTAGAAGATGAATCAGAAATAA 621  
 QY 598 ATGGTAGTTCTTCCAAAGTAAATCTTTGAGAGAAATTTTTCAGCAGATATTACAGATGATATT 657  
 DB 622 AATGTCAATCTCTGGTAAAGCTTTATCTGAAATTAACAAATTAATGATGAGTGAACGAC 681  
 QY 658 GAGACCGTTGAGGTATTTTCTCAACCAAGCAAAATCTTTTTCAGAAAGTGAACACATTTCT 717  
 DB 682 GAAGATATTGATATTTTCTTCTTCTTAAACCAAGTGTATTTCAGAGTGGGAATATTAAAT 741  
 QY 718 TTTTATACACCGCTCTTAGAAGGAATTTATCCCGATACAGACCGTTTATTAATGACAGAA 777  
 DB 742 TTTATCTCACGTTTACTTTGAAGGTCATTTATCCAGATACGACACGTTTATTTCCAGANAAT 801  
 QY 778 TTTGAGACGGAGTTGTTTTCATACCAATCCCTTCCGACGCTTATGGAAGTGCCTTC 837  
 DB 802 TATGAGATTAATTAGGAATTAACATGGAGACTTCTATCATGCAATTTGATGTCATCT 861  
 QY 838 TTGATTTTCAATGCTACTCAAAATGGTACTGTTAAAGCTTGAGATTACTCAAAATCATATT 897  
 DB 862 TTATTAGCACCTGAAGGTGGAATAATGTTTAAATTAAGTACAGGTAATGAATTAGTT 921  
 QY 898 TCAGTCAATGTTAACTCACTGAGGTTGGTAAAGTAAACGAGGATTTAGATATTGTTAGT 957  
 DB 922 GAACCTTTCATCTACTCTCTCTGAAATTTGGTACTGTGTTAAAGAAAGATTTAAACGCTAATGAT 981  
 QY 958 CAGTCTGTAGTGTATTTAACTATCATGCTTCAATCAACTTCACTTATTTGAGTCTTTAAAA 1017  
 DB 982 GTAGAAGCGGAAACTTGAATAATTTCTTCAACTCAAAATACATGATGATGCTTTAAAA 1041  
 QY 1018 GCTATTTAAAGTGAAACAGTAAATAATTCATTTCTTATCACAGTTTCGACCATTCACCTTA 1077  
 DB 1042 GCCATTGATATGATGAAGTAGAAGTAGAATTTCTTTGGTACAATGAAACCATTTATCTTA 1101  
 QY 1078 ACACGACCGATGAGGA 1094  
 DB 1102 AAACCAAAAGATGATGA 1118  
 RESULT 24  
 ACAT47854  
 ID ACA47854 standard; DNA; 1131 BP.  
 XX ACAT47854;  
 XX AC  
 DT 19-JUN-2003 (first entry)  
 XX Prokaryotic essential gene #29511.  
 DE Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KW drug design; gene.  
 XX Staphylococcus haemolyticus.  
 OS WO20027183-A2.  
 XX PD 03-OCT-2002.  
 XX PF 21-MAR-2002; 2002WO-US009107.  
 XX PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zvekind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX

DR WPI; 2003-029926/02.  
DR P-PSDB; ABU43984.

PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
v v

PS Claim 14; SEQ ID NO 35724; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published/pct\\_sequences](http://wipo.int/pub/published/pct_sequences)

Sequence 1131 BP; 403 A; 180 C; 194 G; 354 T; 0 U; 0 Other;

Query Match	21.5%;	Score 243.4;	DB 8;	Length 1131;
Best Local Similarity	52.8%;	Pred. No. 2.4e-42;		
Matches · 598; Conservative	0;	Mismatches 526;	Indels 9;	

Qy 60  
1 ATGATTCAATTTTCAATTAAATCGACATTATTTATTCAGCTTTTAATACAACATAACGT

Db 60  
1 ATGTGGAAATTCACAAATTAGAAGACATTATTTATTAATCAATTAAATGCACATTAAAA

61	GCTATTAGCACTAAAAATGCCATTCCTATTCTTTTCATCAATAAAATTTGAAGTCACTTCT	120
Qy		
61	GCCATCTCCCAAGAACAAACATTACCAATTTTAAAGGGTATCAAAATCGATGCTAAAGAT	120
Db		

QY 121 ACAGGAGTAACTTTAAACAGGGTCTAACCGGTCAAATATCAATTGAAACACTATTCTCTGTA 180

Db 121 AACGAAGTCATTCTTACTGGTTCAGATTCTGAGATATCTATTGAAATTACAAATCCCTAAA 180

181 AGTAATGAAATGCTGGTTTGCTAATTACCTCT---CCAGGAGCTATTTATTAGAAGCT 237  
181 CAAGTAGATGCTGAGGATATTGTCACATATTTCTGAAACAGGTTGAGTTGTACTCTGGT 240

[illegible]

DY  
db

298 GAACAACACCAAGTTCTTTTAAACCAGTCGGTAATCAGNGATTACTCTTAAAAGGAAAAGAT 357  
||||| |||| |  
301 GAACAATTTCAAACACTGATTACTTCAGGCACATCTGAATTTTAACTTAAGTGTTTAGAT 360  
||||| |||| |

358 GTTACCAAGTATCCTCGTCTCAAGAAGTATCAACAGAAATCCTTTCATTAAAAACA 417

361	CTGATCAATATCCAAATTAACCTCAAGTATCACGTGATGATGCAATTAATCAATTAACGTA	420
418	AAATTAATTGAAGTCTATTATTGCTGAAACAGCTTTTGACAGCCAGTTTACAGAAAGTCGT	477
421	AAAGTATTAAAAAATATCATTTGGCACAACAATTTCCGAGTGTCCACCTCAGAAACACGC	480
478	CCATTTTAAACAGGAGTTCAATATGTATTAAGTAATCATAAAGATTTTAAAGCAGTAGCG	537
481	CCAGTACTTACTGGTGTAACTGGCTTATA--CAAGATAATGAATTAATATGCACTGG	537
538	ACTGACTCTCATCTGATCAGCCAAAGTTTAATCACTTTGGACAATCACTTCAGCAGATTG	597
538	ACCGAATCACACCGCTTGGCTGTGAAGAAAGTTAAAAATTAGAAGATGACTCTGAAAAACAA	597
598	ATGGTAGTCTTCCAAGTAAATCTTTGAGAGAAATTTTCAGCAGATATTTACAGATGATATT	657
598	AATGTCAATTCAGGTAAAGCTTTATCAGAAATTAATAAAAAATATGAGTGAAGCGAT	657
658	GAGACCGTTGAGGTATTTTCTCACCAAGCCAAATCTTGTTCAGAAGTGAAACATTTCT	717
658	GAAGAGATTGATATTTCTTTGGCTTCAAAATCAAGTTTATTTAAAGTTGGAATGTAAAC	717
718	TTTTATACACGCTCTTAGAAGGAAATTAATCCGATACAGACCGTTTATTAATCAGACAA	777
718	TTTATTTCTCGCTTATTAGAAGGACATATCTCTGATACAACAGCTTTATTCCTCGTGAAT	777
778	TTTCAGACGGAGTGTGTTTTCAATPACCAATCCCTTCGCCACGCTATGGAACGTCCTTC	837
778	TACGAATAAAAATTAGGCTTAGACAACGGTGAATTTCTATCAGCAATTCGATTCGTCTCT	837
838	TTGATTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTGAGATTACTCAAAATCATATT	897
838	TTATTGGCAGAGAAGGTGGCAACATGTTATTAAAGTTAAGTACGGGTATGATGTTGTA	897
898	TCAGCTCATGTTAACTCACTCAGAGTTGGTAAAGTAAACGAGGATTTAGATATTTGTTAGT	957
898	GAATTAATCATCTACGTCACTGAGATTGGTACTGTAAAGAAGAAGTTACAGCAACGGAT	957
958	CAGTCTGGTAGTGAATTAATCATCAGCTTCAATCCAACTTACCTTATTTAGTCTTTAAAA	1017
958	GTAGAGGCTGGCAACTTGAAGATTTCTTTCACTCTCGTTATATGATGGATTCGGTTAAAA	1017
1018	GCTATTAATAAGTGAAAACAGTAAAAATTCATTTCTTATCACCAAGTTGACCAATTCACCCCTA	1077
1018	GCGATTGATAATGATGAAGTTGAAGTTGAATTTCTTTGGTACGATGAACCGTTCACTCCTT	1077
1078	ACACCGGCGATCAGGAAGAAAGTTTATCCAAATTAATTAACCAAGTACGATGACGAAAC	1130
1078	AAACCA---AAAGAGATGATCTGTCATCTCAATTAATTTTACCAATTAAGAC	1127

RESULT 25

ADEF77343 00

WP Sequence split into 20 fragments LOCUS ADF77343 Accession Adf77343

WP	Fragment Name	Begin	End
----	---------------	-------	-----

WP ADF77343\_00 1 110

WP	ADF77343_01	100001	2100
----	-------------	--------	------

WP	ADF77343_02	200001	3100
----	-------------	--------	------

WP ADF77343\_03 300001 4100

WP ADF77343\_04 400001 510001

WP ADF77343\_05 500001 610001

WF ADF 77343\_06 600001 7:100

WP	ADF 77343_07	7000001	8100
WP	ADF77343_09	9000001	8100

WE	ADE77343-08	800001	9100
WP	ADE77343-09	900001	10100

WP	ADE77343-10	1000001
WZ	ADE77343-09	300001
		I0100

WP ADF77343-11 1100001 1210001

WP ADF77343-12 1200001 1310001 1310001

WP ADF77343-13 1300001 1410001

WP ADF77343-14 1400001 1510001 1510001

WP	ADF77343_15	1500001	161000
----	-------------	---------	--------

WP ADF77343\_16 1600001 171000

WP ADF77343\_17 1700001 181000

WP ADF77343\_18 1800001 1910000  
 ID ADF77343\_19 1900001 1983043  
 XX ADF77343 standard; DNA; 1983043 BP.  
 AC ADF77343;  
 XX  
 DT 26-FEB-2004 (first entry)  
 XX  
 DE Lactic acid bacteria *Lactobacillus johnsonii* Lal genomic DNA SEQ ID NO:1.  
 XX  
 KW lactic acid bacterium; *Lactobacillus johnsonii* strain Lal; detection;  
 XX identification; DNA array; DNA chip; gene; ds.  
 OS *Lactobacillus johnsonii*.  
 XX  
 PN WO2003084989-A2.  
 XX  
 PD 16-OCT-2003.  
 XX  
 PF 19-MAR-2003; 2003WO-EP02882.  
 XX  
 PR 09-APR-2002; 2002EP-00007932.  
 XX  
 PA (NEST ) SOC PROD NESTLE SA.  
 XX  
 PI Pridmore RD, Mollet B, Arigoni F, Hermanns J;  
 XX  
 DR WPI; 2003-804292/75.  
 XX  
 CC Use of a DNA sequence of a *Lactobacillus johnsonii* strain, Lal, or its  
 PT part or sequences homologous to it for elucidating interactions between a  
 PT bacterium and host that it colonizes.  
 XX  
 PS Claim 1; SEQ ID NO 1; 24pp; English.  
 XX  
 CC The present sequence represents a DNA sequence of the lactic acid  
 CC bacteria *Lactobacillus johnsonii* strain Lal. The present invention also  
 CC described the DNA sequence, or its part or sequences homologous to it  
 CC which are useful for elucidating interactions between a bacterium and  
 CC host that it colonizes. Also described: (1) a method for detecting,  
 CC identifying or selecting a *Lactobacillus* strain, preferably *Lactobacillus*  
 CC *johnsonii* in a biological sample; (2) a method for detecting, identifying  
 CC or selecting antibodies directed to *Lactobacillus* polypeptides in a  
 CC biological sample; (3) a DNA array or chip containing an array of  
 CC polynucleotides comprising at least a polynucleotide derived from Lal;  
 CC (4) a protein array or chip containing an array of polypeptides  
 CC comprising at least one of the polypeptides obtainable by expressing a  
 CC polypeptide as identified by an open reading frame derived from Lal; (5)  
 CC one antibody chip containing an array of antibodies comprising at least  
 CC one antibody directed to a polypeptide obtainable by expressing an open  
 CC reading frame in Lal; (6) a screening assay; (7) a kit comprising the  
 CC polynucleotide or antibody raised against the polypeptide; (8) a computer  
 CC readable medium having recorded the nucleic acid sequence of Lal or the  
 CC polypeptide derived from the nucleotide sequence; and (9) a computer-  
 CC based system for identifying fragments of the *Lactobacillus johnsonii*  
 CC genome. The DNA sequence of a *Lactobacillus johnsonii* strain Lal, or its  
 CC part or sequences homologous to it are useful for elucidating  
 CC interactions between a bacterium and host that it colonizes.  
 XX  
 SQ Sequence 1983043 BP; 655813A; 330432C; 354497G; 642272T; 0U; 290ther;  
 Query Match 21.3%; Score 242; DB 10; Length 110000;  
 Best Local Similarity 53.4%; Pred. No. 1e-41;  
 Matches 609; Conservative 0; Mismatches 510; Indels 21; Gaps 4;  
 QY 1 ATGATTCAATTTTCAATTAATCGCATATTTATTTATTCATGCTTTTAAATACAACTAAACGT 60  
 DB 1543 ACGATGCGAGTTACAAATTAATCGTATTTATTTCTCGAAACCTAAATTAATGCTATGCGT 1602  
 QY 61 GCTATTAGCATTAAATGCAATTCCTATTTCTTCATCAATATAAAATGAAGTCACTTCT 120  
 DB 1603 GCAATTTCTTACGGTCTACTATTCCATATTAAGTGGTATTAATACTTACCCTTACAGT 1662

QY 121 ACAGAGTAACCTTTAAACAGGGTCTTAACGGTCAAAATATCAATTTGAAAAACACATTTCTCTGTA 180  
 DB 1663 GAAATGCTTACTTTAAACCGGTAGTACTGACATTTCTTATTGAAATTTCAATCCCACT- 1721  
 QY 181 AGTAATGAAATGCTGGTTTCTGCTAAATACCTCTCCAGGAGCTATTTTATTAGAGCTAGT 240  
 DB 1722 -----AAACGATGATTTAATTTGTTCAATCTACGGATCGATCGTTTTTACCAGACGA 1773  
 QY 241 TTTTATTAATATTTTCAAGTTTGCAGATATTTAGTATATAAATGTTTAAAGAAATTTGAA 300  
 DB 1774 TTTTTCAGTGAATTTGAAAAAATTTACCAGTAAAGATTTTCTATTGAAAGTAAAGAA 1833  
 QY 301 CAACACCAAGTTGTTTAAACAGTGTGTAATCAGAGATTACCTTTAAAGGAAAAAGATGTT 360  
 DB 1834 AGTTTTCAAACGAAAAATTTGTTTCTGAAATACCGAATTCATGATTAACGGTTTAGATGA 1893  
 QY 361 GACCATATCTCTGCTACAAAGAGTATCAACAGAAAAATCTTTGATTTTAAAAACAAAA 420  
 DB 1894 AACAAATATCCACACTTACAGAAATTTCTACTGATGATCATTTCAAAATTTCTGGTAAA 1953  
 QY 421 TTATTGAAGTCTATTTATTTGCTGAACAGCTTTTGCAGCCAGTTTCAAGAAAGTCTCT 480  
 DB 1954 ACATTTAGAGAAATTTATTAACGAAATCTGTTTTTGCAGTCGCTACTCAAGAAAGTCTCCA 2013  
 QY 481 ATTTTAAACAGAGTCTATTTGTTAAGTAAATCATAAAGATTTTAAAGCAGTAGCGACT 540  
 DB 2014 ACTTTAACTGGTGAATCTTTATCTT---CAACAATTCATCAATTTAAAGCAGTTGCTACC 2070  
 QY 541 GACTCTCATCTGATGAGCCAACTTTAATCACTTTGGACAATA---CTTCAGCAGATTG 597  
 DB 2071 GACAGTACCAGTATCTCAACGTCMAATTTCTTAGAAATGGTCTTCAACAGTACT 2130  
 QY 598 ATGGTAGTTCTTCCAAGTAAATCTTTGAGAGAAATTTTTCAGCAGTATTTTACAGATGATAT 657  
 DB 2131 GACTTAATTTATCTCGGAAAGAGCTTAGTAGAATTTATCCCGAATTTATCGGAGAAAGTAT 2190  
 QY 558 GAGACGGTGGAGTATTTTCTCCACCAAGCAAAATCTTTGTCAGNAGTGAACACATTTCT 717  
 DB 2191 CCTGAAATCACAGTAAATCTGGTGGAAACCAAGTTTATTTGAAAGTTGGAACAACTGCA 2250  
 QY 718 TTTTATACACCCCTCTTTAGAGAGAAATTTATCCCATACAGACCGCTTTTATTAATGACAGAA 777  
 DB 2251 TTCTATTACGTTTACTTGTAGTGGACATATCCAGACACTGATCGTTTAAATTTCCAATGAA 2310  
 QY 778 TTTGAGACGGAGTGTGTTTCAATACCAATCCCTTCGCCACGCTATGGAACGTCCTTC 837  
 DB 2311 TCTACAACTCAGTTGAATTTTGAATTTACAGTTTGTAGCTCTCTCTCGAACGTCGCACT 2370  
 QY 838 TTGATTTCTAATGCTACTCAAAATGGTACTGTTAAG-----CTTGAGATTTACTCAAAAT 891  
 DB 2371 CTTCTTACTACGAAAGCGGTAATTAATGTTGTTAAGATGACTCTTTGATTTTCAAAATCAA 2430  
 QY 892 CATATTTTCAGCTCATGTTAACTCACTCAGGTGTTGTAAGGTAAGCAAGGATTTAGATATT 951  
 DB 2431 TTAGTTAACTCCAGGTGATTTACCGGAATAGGAATAGGAATGGAAGAAATTTGGCTTT 2490  
 QY 952 GTTAGTCTAGTCTGGTAGTATTTAACTATCAGCTTTCAATCCAACTTACTTATTGAGTCT 1011  
 DB 2491 AAGAATCTTGAAGGAGATGGCTTAATTTCTTCAATCTCTGATTTATTAAGAGAGCT 2550  
 QY 1012 TTAAGAGTATTAAGAGTGAACAGTAAATTTCTTTATCATCCAGTTTCGACCATTC 1071  
 DB 2551 TTACGTGCTCTATTACTGATTTCCATTTATTAAGTAACTTTACCCAGCCGCTAAGACATTT 2610  
 QY 1072 ACCCTAACACACGAGCGATGAGAGAGAAAGTTTATCCAAATTAATTAACACAGTACGAACA 1131  
 DB 2611 ACAGTGATACCGCAACAAAGATGTTAATTTCTCAATTAATCAACGCCAGTTAGAAC 2670  
 RESULT 26  
 ACF74365  
 ID ACF74365 standard; DNA; 1131 BP.  
 XX











PR 04-OCT-2001; 2001US-0327193P.  
 PR 30-OCT-2001; 2001US-0340922P.  
 PR 05-NOV-2001; 2001US-0338709P.  
 PR 06-NOV-2001; 2001US-0333269P.  
 PR 18-DEC-2001; 2001US-0341679P.  
 XX  
 PA (AFPI-) AFFINIUM PHARM INC.  
 XX  
 XX Edwards A, Dharamsi A, Vedadi M, Alam MZ, Arrowsmith C, Awrey D;  
 PI Beattie B, Canadian V, Cox B, Domagala M, Houston S, Li Q;  
 PI Netherly K, Ng I, Ouyang H, Pinder B, Sheldrick B, Viola C;  
 PI Wrezel O;  
 XX  
 XX WPI; 2003-371793/35.  
 DR P-PSDB; ADD26272.  
 XX  
 PT New crystallized recombinant polypeptides from *Staphylococcus aureus*,  
 PT *Streptococcus pneumoniae*, *Helicobacter pylori*, *Pseudomonas aeruginosa*  
 PT involved in nucleic acid processing, useful as targets for pathogenic  
 PT bacteria.  
 XX  
 PS Claim 218; SEQ ID NO 76; 298pp; English.  
 XX  
 CC The present invention describes a crystallised recombinant polypeptide  
 CC (I) comprising the amino acid sequence of polypeptides from  
 CC *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Helicobacter pylori* and  
 CC *Pseudomonas aeruginosa* which are involved in nucleic acid processing, or  
 CC amino acid sequences having at least 90 % identity with the polypeptide  
 CC sequence, where the polypeptide is in crystal form. (I) comprises the  
 CC amino acid sequence (S) of polypeptides involved in nucleic acid  
 CC processing, which includes ribonuclease diphosphate reductase, major  
 CC subunit (nrDE), polyribonucleotide phosphorylase (pnPA), transcription  
 CC termination factor Rho (Rho), putative polynucleotide  
 CC nucleotidyltransferase (pnp), DNA ligase (lig), DNA polymerase III, beta-  
 CC subunit (dnan) or ribonucleoside diphosphate reductase minor subunit (R2)  
 CC (nrDF) from *S. aureus*; uridylylate kinase (pyrH) or orotate  
 CC phosphoribosyltransferase (pyrE) from *H. pylori*; uracil-DNA-glycosylase  
 CC (lung) from *P. aeruginosa*; and nrDE from *S. pneumoniae*. (I) is in a  
 CC crystal form. (I) has antibacterial activity, and can be used in  
 CC vaccines. (I) is useful for designing a modulator for the prevention or  
 CC treatment of *S. aureus*, *S. pneumoniae*, *H. pylori* and *P. aeruginosa*  
 CC related diseases or disorders. The method comprises providing a three-  
 CC dimensional (3D) structure for (I), identifying a potential modulator by  
 CC reference to the 3D structure, contacting the potential modulator with  
 CC the recombinant polypeptide and assaying the activity of the polypeptide  
 CC or determining the viability of *S. aureus*, *S. pneumoniae*, *H. pylori* or *P.*  
 CC *aeruginosa* after contact with the modulator, where a change in the  
 CC activity of the polypeptide or the viability of the bacteria indicates  
 CC that the modulator may be useful for preventing or treating the disease  
 CC or disorder. The structural and functional information of (I) aid in the  
 CC discovery and design of therapeutic and diagnostic molecules. The present  
 CC sequence is used in the exemplification of the present invention.  
 XX  
 SQ Sequence 1134 BP; 410 A; 177 C; 195 G; 352 T; 0 U; 0 Other;

Query Match 20.8%; Score 236.2; DB 10; Length 1134;

Best Local Similarity 52.2%; Pred. No. 8.1e-41;

Matches 573; Conservative 0; Mismatches 518; Indels 6; Gaps 2;

QY 1 ATGATTCAATTTTCAATTAATCGCATTATTTATTTATTCATCAATAAATAATGAAGTCACTTCT 60  
 DB 1 ATGATGGAAATTCACATTAATAAGAGATTATTTATTTATACAAATTAATGACATTAATAA 60  
 QY 61 GCTATTAGCAGCTAAATAATGCCATTCTCTATTTCTTCATCAATAAATAATGAAGTCACTTCT 120  
 DB 61 GCTATTTCACCAAGAACCAATACCTATATTAATTAATCGGTATCAAAATCGATGCGAAGAA 120  
 QY 121 ACAGAGATCACTTTAAACGGGGCTAACCGGTCAAAATATCAATTAATGAATAATCTTCTGTA 180  
 DB 121 CATGAAGTTATTAATCACTGGTTTCAGACTCTGAAATTTCAATAGAATAATCACTATTCTTAA 180  
 QY 181 AGT---AATGAAATGCTGGTTGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 237

DB 181 ACTGTAGATGCGGAAGATATTGTCAATATTTCAGAAAACAGGCTCAGTAGTACTTCTCTGA 240  
 QY 238 AGTTTTTTAAATTAATTAATTTCAAGTTTGGCAGATATTAGTATATAAATGTTAAAGAAATTT 297  
 DB 241 CGATCTTTTGTGATATTATAAATAAATAATTAAGTAAAGATGTTAAATATTATCTACAAT 300  
 QY 298 GAACAAACCAAGTTGTTTAAACGAGTGTAAATCAGAGATTACCTTTAAAGGAAAAGAT 357  
 DB 301 GAACAAATCCAGACATTAAATTAATCAATCAGGTCAATCTGAAATTTAAATTAAGTGGCTTAGAT 360  
 QY 358 GTTGACCAAGTATCTCGCTACAGAGAGTATCAACAGAAAATCCCTTTGATTTTAAACACA 417  
 DB 361 CCAGATCAATATCTTTTATTAATCTCAAGTTCTTAGAGATGACGCAATCAATTTGTCGTA 420  
 QY 418 AAATTAATGAAGTCTATTATTGCTGAAACAGCTTTTGCAGCCAGTTTACAAGAAAAGTCTG 477  
 DB 421 AAAGTGTCTTAAACAGCTGATTGCAACAAATAATTTGCAAGTGTCCACCTCAGAAAACAGC 480  
 QY 478 CTTATTTTAAACAGGAGTTCATATTGTTAGTAAATCATTAAGATTTTAAAGCAGTAGCG 537  
 DB 481 CCAGTACTAACTGGTGAACCTGGCTTATA---CAAGAAAATGAATTAATATGCAACGCG 537  
 QY 538 ACTGACTCTCATCGTATGAGCAACAGTTTAAATCACTTTGGCAATATCTTCAGCAGATTG 597  
 DB 538 ACTGACTCACACCGCTGGCTGTAGAAGTTGCGTTAGAAGATGTTTCTGAAAACAAA 597  
 QY 598 ATGTGAGTTCTTCCAAGTAAATCTTTGAGAGAAATTTTCCAGCAGTATTTACAGATGATATT 657  
 DB 598 AATGTGATCAATCCAGGTGAAGCTTTTAGCTGAATTAATAAATAATTTATGCTGACAATGAA 657  
 QY 658 GAGACCGTTGAGTATTTTCTTCAACAGCAAAATCTTGTTCAGAAAGTGAACACATTTCT 717  
 DB 658 GAAGACATTTGATATCTTCTTTGCTTCAACCAAGTTTATTTTAAAGTTGGAATGTAAC 717  
 QY 718 TTTTATACACCGCTCTTAGAAGGAAATTTATCCGATACAGACCGTTTATTAATTAATGACAGAA 777  
 DB 718 TTTATTTCTCGATTATTAAGAAGCAATTTCTGATACACACCGTTTATTTCCCTGAAAAC 777  
 QY 778 TTTGAGACGGAGTTGTTTTCATACCAATCCCTTCCACCGCTATGGAACGTCCTTC 837  
 DB 778 TATGAAATTAATAATTAAGTATAGACAAATGGGAGTTTATCATGCGATTGATCGTCCCTCT 837  
 QY 838 TTGATTTCTAATGCTACTCAAAATGGTACTGTTAGCTTTGAGATTACTCAAAATCATATT 897  
 DB 838 TTATTAGCGCTGAAGGTGGTAATAACGTTTATTAATTAAGTACAGGTGATGACGTTGTT 897  
 QY 898 TCAGCTCATGTTAACTCACTGAGCTTGGTAAGGTAAACGAGGATTTAGATATTCTTAGT 957  
 DB 898 GAATTTGCTTCTACATCACAGAAATTTGTTACTGTAAAGAAGAAGTTGATGCAAAACGAT 957  
 QY 958 CAGTCTGTGAGTATTTAACTTATCAGCTTCAATCCAACTTACCTTTATTTAGTCTTTTAA 1017  
 DB 958 GTTGAAGGTGTAGCTGCTGAAAATTTTCAATCAACTCTAAATATATGATGATGCTTTTAAA 1017  
 QY 1018 GCTATTAAGTGAACAGTAAATAATTCATTTCTTATCAGGTTGACCACTTCGACCATTCACCTA 1077  
 DB 1018 GCAATCGATAATGATGAGGTTGAAAGTTGAATTTCTTCGTTACAATGAACCAATTTATCTTA 1077  
 QY 1078 ACACGAGCGATGAGGA 1094  
 DB 1078 AAACCAAAAGGTGAGGA 1094  
 RESULT 30  
 ADD26273  
 ID ADD26273 standard; DNA; 1134 BP.  
 XX  
 AC ADD26273;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE *Staphylococcus aureus* dnaN DNA SEQ ID NO:78.  
 XX



Db 898 GAATGCTCTTACATCACCAGAAATGGTACTGTAAAGAGAGAGTGTGATGCAACAGAT 957  
 Qy 958 CAGTCTGGTAGTGAATTAACATATCAGCTTCAATCAACTTACCTTAATGAGCTTTAAAA 1017  
 Db 958 GTTGAAGGTGGTAGCTGAAATTTCAATCAACTCTAAATATATATGATGATGCTTTAAAA 1017  
 Qy 1018 GCTATTAAAGTGAACAGTAAATTCATTTCTTATCACCAGTTCGACCAATTCACCTA 1077  
 Db 1018 GCAATCGATATGATGAGGTTGAAGTTGAATTTCTTCGATACAAATGAACCACTTATTCTA 1077  
 Qy 1078 ACACAGGCGATGAGGA 1094  
 Db 1078 AACCAAAAGGTGACGA 1094

RESULT 31

AAV74464

ID AAV74464 standard; DNA; 2347 BP.

XX AC

XX AAV74464;

XX 16-MAR-1999 (first entry)

XX Staphylococcus aureus contig SEQ ID #153.

XX Computer readable medium; vaccine; S.aureus infection; immunodetection;

XX cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

XX skin infection; surgical wound infection; scalded skin syndrome;

XX toxic shock syndrome; ds.

XX Staphylococcus aureus.

XX Key Location/Qualifiers

FT 1741..1800

FT misc\_feature

FT /tag= a

FT /note= "these bases represent a line of missing text in

FT the sequence listing in the specification. They are

FT included to maintain the nucleotide numbering given in

FT the specification for this DNA sequence"

XX EP786519-A2.

XX 30-JUL-1997.

XX 07-JAN-1997; 97EP-00100117.

XX 05-JAN-1996; 96US-0009861P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Kunsch CA, Choi GH, Barash SC, Dillon PJ, Fannon MR, Rosen CA;

XX WPI; 1997-374922/35.

XX Polynucleotide(s) and proteins derived from Staphylococcus aureus -

XX stored on computer readable medium and used in the production of anti-

XX S.aureus vaccines.

XX Claim 1; Page 775-776; 3271pp; English.

XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences

XX of the invention. The DNA sequences are recorded on a computer readable

XX medium, preferably selected from a floppy or hard disk, random access

XX memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using

XX the S.aureus DNA sequences allows putative functions to be assigned so

XX that protein-encoding or regulatory regions of commercial, therapeutic or

XX industrial importance can be obtained. Specifically, sequences which are

XX likely to encode antigens have been identified and these polypeptides can

XX be used in a vaccine composition against S.aureus infection. The

XX polypeptides can also be used in a kit for the immunodetection of

XX S.aureus in a sample. S.aureus is implicated in numerous human diseases,

XX including cellulitis, eyelid infections, food poisoning, osteomyelitis,

CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S.aureus DNA sequences contained on the computer  
 CC readable medium

XX Sequence 2347 BP; 865 A; 331 C; 382 G; 707 T; 0 U; 62 Other;

Query Match 20.8%; Score 236.2; DB 2; Length 2347;

Best Local Similarity 52.2%; Pred. No. 9.2e-41;

Matches 573; Conservative 0; Mismatches 518; Indels 6; Gaps 2;

Qy 1 ATGATTCAATTTTCAATTAATCGCACATATTATTATTCATGCTTTTAAATACAACCTAAACGT 60

Db 393 ATGATGGAATTCACCTATTAAAAGAGATTTATTTATACACAATTTAAATGACACATTAAAA 452

Qy 61 GCTATTAGCACTAAATGCGCATTCCTATTCTTTTCATCAATAAATAATGAGTCACTTCT 120

Db 453 GCTATTTCCCAAGAACCAACATTACCTATATTAATCTGGTATCAAAATCGATCGGAAGAA 512

Qy 121 ACAGGAGTAACTTTAAACAGGGGTCTAACGGTCAAAATATCAATTTGAAAAACACTATTCTGTA 180

Db 513 CATGAAGTTATATTAACCTGTTTCAGACTCTGAAATTTCAATAGAAATCACTATTCTTAAA 572

Qy 181 AGT---AATGAAATGCTGGTTGCTAATTAACCTCTCCAGGAGCTATTTTATTAGAAGCT 237

Db 573 ACTGTAGATGGCGAGATATTGTCAATATTTTCAAGAACAGGCTCAGTAGTACTTCTGGA 632

Qy 238 AGTGTATTTTAAATATTTTCAAGTTTCCAGATATTAGTATATAATGTTTAAAGAAAT 297

Db 633 CGATTTCTTTGTTGATATTATAAAAAAATTACCTGTTAAAGATGTTTAAATTTATCTCAAA 692

Qy 298 GAAACAAACCAAGTTGTTTAAACAGGTGCTGTAATCAGAGATTACCTTTAAAGGAAAAAGAT 357

Db 693 GAACTATTCAGACATTAATTTACATCAGCTCATCTGAAATTTAAATTTAAGTGGCTTAGAT 752

Qy 358 GTTGACAGTATCTCTGCTCAAGAGATATCAACAGAAAAATCTTTGATTTTAAAAACA 417

Db 753 CCAGATCAATATCTCTTTATTACCTCAAGTTTCTAGAGATGACGCAATTTCAATTTGCGGTA 812

Qy 418 AAATTTTGAAGTCTATTATTGCTGAAACAGCTTTTGCGAGCGAGTTTACAGAAAGTCTG 477

Db 813 AAAGTCTTAAACAGCTGATTTGCAACAAACAAATTTTGCGAGTGCACCTCAGAACACCC 872

Qy 478 CCTATTTTAAACAGGAGTTTCATATTGTTAAGTATATCAATAAGATTTTAAAGCAGTAGCG 537

Db 873 CCAGTACTAACTGGTGTGAACCTGGCTTATA---CAAGAAATGAATTAATATGACAGCG 929

Qy 538 ACTGACTCTCATCGTATGAGCCAAAGCTTTAATCACTTTTGGCAAAATCTTTACAGAGATTG 597

Db 930 ACTGACTCACACCGCTTGGCTGTAAGAAAGTTGCAAGTTAGAGATGTTTCTGAAAAACAA 989

Qy 598 ATGTTAGTCTTCCAAAGTAATCTTTGAGAGAAATTTTCAGCAGTATTACAGATGATATT 657

Db 980 AATGTCATCTCCAGGTAAGCTTTAGCTGAAATTAATAATAAATTTATGTCTGACAAATGAA 1049

Qy 658 GAGACCGTTGAGGTATTTTCTCTCAACAAAGCAAAATCTTTGTTGAGAGTGAACACATTTCT 717

Db 1050 GAAGACATTTGATATCTTCTTCTTCAACCCAAAGTTTATTAAAGTTGGAATTTGTAAC 1109

Qy 718 TTTTATACACGCTCTTAGAAGGAAATTTATCCGATACAGACCGTTTATTATATGACAGAA 777

Db 1110 TTTATTTCTCGATTATTAGAAGGACATTTATCTGTATACAAACGCTTTTATTTCCCTGAAAAC 1169

Qy 778 TTTGAGCGGAGGTGTTTTCATACCAATCCCTTCGCCAGCTTATGGAAGCTGCTTC 837

Db 1170 TATGAAATTAATTAAGTATGACAAATGGGAGTTTATCATGCGGATGATCGTGCCTCT 1229

Qy 838 TTGATTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTTGAGATTACTCAAAATCATATT 897

Db 1230 TTATTAGCGGTGAAGGTGGTAAATTAACGTTATTAAATTAAGTACAGGTGATGACGTTGTT 1289

QY 898 TCAGTCATGTTAACTACCTGAGGTTGGTAAAGTAACGAGGATTTAGATTTGTTAGT 957  
 Db 1290 GAATTTGCTCTTACATACACGAAATTTGTTAAAGAAAGTTGATGCAACGAT 1349  
 QY 958 CAGTCTGTAGTATTAACATATCAGCTTCAATCACTTACCTTATTGAGTCTTTAAAA 1017  
 Db 1350 GTTGAAGGTGTAGCTGAAATTTCACTCACTAAATATATGATGATGCTTTAAAA 1409  
 QY 1018 GCTATTTAAAGTGAACAGTAAATATCTTCTATCACCAGTTCACCACTTACCCCTA 1077  
 Db 1410 GCAATCGATATGATGAGTTGAATTTCTTGGTCAATGAACCAATTTATTTCTA 1469  
 QY 1078 ACACAGCGGATGAGGA 1094  
 Db 1470 AAACCAAAAGGTGACGA 1486

RESULT 32

AAZ31005

ID AAZ31005 standard; DNA; 1134 BP.

AC AAZ31005;

XX 05-JAN-2000 (first entry)

DT Partial dnaN gene.

DE Gram positive bacteria; dnaE; dnaX; dnaB; dnaG; dnaN; dnaG; helicase;

KW alpha subunit; DNA polymerase III holoenzyme; gamma subunit; tau subunit;  
 KW clamp loader; glue protein; replication; antibiotic; ss.  
 XX Staphylococcus aureus.

OS

XX

XX

Key Location/Qualifiers

FT CDS 1..1134

FT /\*tag= a

FT /note= "Dna N gene product"

XX WO9937661-A1.

XX 29-JUL-1999.

XX 25-JAN-1999; 99WO-US001547.

XX 27-JAN-1998; 98US-0074522P.

XX 22-JUL-1998; 98US-0093727P.

XX (UYRQ ) UNIV ROCKEFELLER.

XX O'donnell ME, Zhang D, Whipple R;

XX WPI; 1999-590685/50.

XX P-PSDB; AAY49071.

XX New isolated dnaE, dnaX and dnaB genes from Gram positive bacteria, used  
 to develop screening assays for identifying antibiotic compounds.  
 Example 8; Page 30; 132pp; English.

XX This sequence is the partial dnaN gene of *Staphylococcus aureus*. The  
 CC invention relates to a number of isolated DNA molecules from Gram  
 CC positive bacterium, corresponding to dnaE (AAZ31001), dnaX (AAZ31002), and  
 CC dnaB (AAZ31003). The PolC, dnaN and dnaG genes (AAZ31004-231006) are also  
 CC identified. The dnaE gene corresponds to the alpha subunit of the  
 CC Escherichia coli, DNA polymerase III holoenzyme, dnaX corresponds to the  
 CC gamma and tau subunits, and dnaB corresponds to the helicase. The alpha  
 CC subunit is the actual DNA polymerase, the gamma complex forms the clamp  
 CC loader and tau is a "glue protein". dnaX encodes both gamma and tau, Tau  
 CC is the product of the full gene, while gamma is the product of the first  
 CC third of the gene. dnaN forms the beta subunit which forms the  
 CC sliding clamp, and dnaG encodes a primase. The DNA sequences of the  
 CC invention can be used to identify agents that inhibit or promote DNA  
 CC replication by acting on various parts of the gram positive bacterial DNA

CC polymerase holoenzyme. The products and methods of the invention can be  
 CC used for identifying pharmacological agents or lead compounds for agents  
 CC active at the level of a replication protein function, particularly DNA  
 CC replication. The agents identified can be used as antibiotics  
 XX

SQ Sequence 1134 BP; 411 A; 176 C; 194 G; 353 T; 0 U; 0 Other;

Query Match 20.5%; Score 233; DB 2; Length 1134;

Best Local Similarity 52.1%; Pred No. 3.9e-40;

Matches 571; Conservative 0; Mismatches 520; Indels 6; Gaps 2;

QY 1 ATGATTCATTTTCAATTAATGCGACATTTATTTATTCATGCTTTTAAATFACAACTAAACGT 60

Db 1 ATGATGGAATTCACATTTAAAGAGATTTATTTATACACAATTAATGACACATTTAAA 60

QY 61 GCTATTAGCAGCTAAATGCGCATTCCTTCTTCATCAATAAAATGAAGTCACTTCT 120

Db 61 GCTATTTCCACCAAGAACCAATTTACCTATATTAATGATCAAAATCGATGCGAAAGAA 120

QY 121 ACAGAGTAATTTTAAACAGGCTCTAACGGTCAAAATATCAATTTGAAACACATTTCTCTGTA 180

Db 121 CATGAAGTTATTAATTAAGTCTGAGACTCTGAAATTTCAATAGAAATCACTATTTCTTAAA 180

QY 181 AGT---AATGAAATGCTGTTGCTAATTAACCTCTCCAGAGCTATTTTATAGAGCT 237

Db 181 ACTGTAGATGCGAAGATATTTGCAATTTTCAGAAACAGGCTCAGTAGTACTTCTCTGGA 240

QY 238 AGTTTTTTTATTAATTTTCAAGTTTCCAGATATTTAGTATATAATTTTAAAGAAAT 297

Db 241 CGATTTCTTTGTTGATATTAATAAAATTAACCTGGTAAAGATTTAAATTTATCTACAAAT 300

QY 298 GAACAACACCAAGTTGTTTAAACAGTGGTAATTCAGAGATTTACCTTTAAAGGAAAGAT 357

Db 301 GAACAATTCAGACATTAATTAATCAATGAGTCACTCTGAATTTAATTTAGTGGCTTAGAT 360

QY 358 GTTGACCAAGTATCCTGCTCTACAAGAGATATCAACAGAAATCTTTGATTTTAAAAACA 417

Db 361 CCAGATCAATATCTTTATTTACCTCAAGTTTCTAGAGATGACCAATTCATTTGTCGGTA 420

QY 418 AATATTGAAGTCTATTTATGCTGAAACAGCTTTTTCAGCCAGTTTCAAGAAAGTCTGT 477

Db 421 AAAGTACTTAAACCGTGTGACAAACGAAATTTTGCAGTGTCCACCTCAGAAACACGC 480

QY 478 CCTATTTTAAACAGGATTCATATTTGTTAAAGTAATCATAAAGATTTTAAAGCAGTAGCG 537

Db 481 CCAGTACTAATGCTGTGACTGGCTTATA---CAAGAAATGAATTAATATATGCACAGCG 537

QY 538 ACTGACTCTCATGTATGAGCCAAACGTTTAAATCACTTTTGGCAATATCTTTTTCAGCAGATTG 597

Db 538 ACTGATTTCAACCGCTTGGCTGTAAAGAGTTTGCAGTTAGAGATGTTTCTGAAACAA 597

QY 598 ATGATAGTTTCTTCCAAAGTAATTTTTCAGAGAAATTTTCAGCAGTATTTTACAGATGATTT 657

Db 598 AATGTCATCTTCCAGGTAAGGCTTTAGCTGAATTTAAATTAATTTATGCTGACAAATGAA 657

QY 658 GAGACCGTTGAGGTATTTTCTCACCAGCCAAATCTTGTTCAGAGAGTGAACACATTTTCT 717

Db 658 GAAGACATTTGATATCTTTCTTCTTCAACCAAGTTTATTTTAAAGTGGAAATGTGAAC 717

QY 718 TTTTATACAGCTCTTACAGAGAAATTTTCCGATACAGCCGTTTATTTAATAGACAA 777

Db 718 TTTATTTCTCGATTTATTAAGAGGACATTTCTGTAACAACACGTTTATTTCTTCTGAAAC 777

QY 778 TTTGAGACGGAGTTGTTTTTCAATACCAATCCCTTCCGAGCTTATGGAAGTCCCTTTC 837

Db 778 TATGAAATTAATTAAGTATAGCAATGGGAGTTTATCATGCAATTTGATGCTGCTCT 837

QY 838 TTGATTTCTAATGCTACTCAAAATGCTTACTGTTAGCTTGAGATTTACTCAAAATCATATT 897

Db 838 TTATTAGCAGCGGAAGGTGGTAAATTAACGTTTATTAATTAAGTACAGGTGATGACCTGTT 897

QY 898 TCAGCTCATGTTAACTCACCTGAGGTGCTGAAGGTAAACGAGGATTTAGATTTGTTAGT 957

Db 898 GAATTATCTTCTACATCACCAGAAATTTGGTACTGTAAAGAAAGAGTTGATGCAAAACGAT 957  
 QY 958 CAGCTCGTGGTAGTATTAACTATCAGCTTCAATCCAACTTACCTTATTGAGTCTTTAAAA 1017  
 Db 958 GTTGAAGGTGGTAGCTGAAATTTCAITCAACTCTAAATATATGATGATGCTTTAAAA 1017  
 QY 1018 GCTATTAAAGTGAACAGTAAATTCATTTCTTATCACAGTTCGACCATTCACCTTA 1077  
 Db 1018 GCAATCGATATGATGAGGTTGAAGTTGAATTTCTTCGGTACAAATGAAACCATTTATTCTA 1077  
 QY 1078 ACACGAGCGATGAGGA 1094  
 Db 1078 AAACCAAAAGGTGACGA 1094  
 RESULT 33  
 AAS54965  
 ID AAS54965 standard; DNA; 1134 BP.  
 XX AAS54965;  
 AC AAS54965;  
 XX  
 XX 13-FEB-2002 (first entry)  
 XX  
 XX Staphylococcus aureus DNA for cellular proliferation protein #1277.  
 DE Staphylococcus aureus DNA for cellular proliferation protein #1277.  
 XX  
 KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;  
 KW antibacterial; drug design.  
 XX  
 XX Staphylococcus aureus.  
 OS  
 XX  
 XX WO200170955-A2.  
 PN  
 XX  
 XX 27-SEP-2001.  
 XX  
 XX 21-MAR-2001; 2001WO-US009180.  
 PP  
 XX  
 XX 21-MAR-2000; 2000US-0191078P.  
 PR 23-MAY-2000; 2000US-0206848P.  
 PR 26-MAY-2000; 2000US-0207727P.  
 PR 23-OCT-2000; 2000US-0242578P.  
 PR 27-NOV-2000; 2000US-0253625P.  
 PR 22-DEC-2000; 2000US-0257931P.  
 PR 16-FEB-2001; 2001US-0269308P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 PI  
 XX  
 DR WPI; 2001-611495/70.  
 XX P-PSDB; AAU37106.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids.  
 XX  
 PS Claim 27; SEQ ID NO 8602; 511bp; English.  
 XX  
 XX The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the genes,  
 CC their use in the discovery of novel antibiotics, the essential genes,  
 CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
 CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
 CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
 CC useful for the identification of potential new targets for antibiotic  
 CC development. The antisense nucleic acids can also be used to identify  
 CC proteins used in proliferation, to express these proteins, and to obtain  
 CC antibodies capable of binding to the expressed proteins. The proteins can  
 CC be used to screen compounds in rational drug discovery programmes. The  
 CC antisense nucleic acid sequence is also useful to screen for homologous  
 CC nucleic acids which are required for cell proliferation in a wide variety  
 CC of organisms. The present sequence encodes an essential prokaryotic  
 CC cellular proliferation protein. Note: The sequence data for this patent  
 CC did not form part of the printed specification, but was obtained in

CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 1134 BP; 414 A; 175 C; 192 G; 353 T; 0 U; 0 Other;  
 Query Match 20.5%; Score 233; DB 4; Length 1134;  
 Best Local Similarity 52.1%; Pred. No. 3.9e-40;  
 Matches 571; Conservative 0; Mismatches 520; Indels 6; Gaps 2;  
 QY 1 ATGATTCAAATTTTCAATTAATGCGACATTAATTTATTTATTTTAAATACAACTAAACGT 60  
 Db 1 ATGATGGAAATTCACATTTAAAGAGATTTATTTATACAAATTAATGACACATTAAAA 60  
 QY 61 GCTATTAGCACAATAAATGCCATTCTCTATTCTTTTCATCAATAAATAATTTGAAGTCATCTCT 120  
 Db 61 GCTATTTTCAACAAGAACCAACATTACCTATATACTGGTATTAATAATCGATGCAAAAGAA 120  
 QY 121 ACAGGAGTAACCTTTAAACAGGCTTAACGGTCAATATCAATTAATGAAAACACTATTCTCTGA 180  
 Db 121 CATGAAGTTATTAACCTGGTTTCAGACTCTGAAATTTTCAATAGAAATCACTATTCTCTAAA 180  
 QY 181 AGT---AATGAAAATGCTGGTTTGTCTAATTAACCTCTCCAGGAGCTATTATTTATAGAAGCT 237  
 Db 181 ACTGTAGATGCGAGATATTGTCAATATTTTCAAGAACAGGCTCAGTAGTACTTCTCTGGA 240  
 QY 238 AGTATTTTAAATATTAATTTCAAGTTTGCAGATATTAGTATAAATGTTTAAAGAAATTT 297  
 Db 241 CGATTTCTTTGTGATATTATAAAAAAATTACCTGCTAAAGATGTTAAATTTATCTACAAAT 300  
 QY 298 GAACAACACCAAGTTGTTTAAACCAGTGGTAAATCAGAGATTAACCTTTAAAGGAAAGAT 357  
 Db 301 GAACAATTCACAGACATTAATTAACATCAGGCTCAATCTGAAATTTAAATTTAAGTGGCTTTAG 360  
 QY 358 GTTGACAGATATCTCGTCTACAAGAGTATCAACAGAAAAATCCTTTGATTTTAAAAACA 417  
 Db 361 CCAGATCAATATCTTTATTAACCTCAAGTTTCTAGAGATGAGCAATTTCAATTTGTCGTA 420  
 QY 418 AAATTTATTGAAGTCTATTATTGCTGAAACAGCTTTTGCAGCCAGTTTACAGAAAGTGGT 477  
 Db 421 AAAGTGCTTTAAAAACGTAATTTGCAAAACGAATTTTGCAGTGTCCACCTCAGAAAACAGC 480  
 QY 478 CCTATTTTAAACAGGATTCATATTGTATTAGTAAATCAATAAGATTTTAAAGCAGTAGCG 537  
 Db 481 CCAGTACTACTGTTGGTGAACCTGGCTTATA---CAAGAAAATGAATTAATATGCAACGG 537  
 QY 538 ACTGACTCTCATCGTATGAGCAACGTTTAACTTTGGACAATACATTTTACAGCAGATTG 597  
 Db 538 ACTGACTCACACCGCTTGGCTGTAAAGAAAGTTGCAGTTAGAGATGTTTCTGAAAACAAA 597  
 QY 598 ATGAGTAGTTCTTCCAAGTAAATCTTTGAGAGAAATTTTCAGCAGTATTTTACAGATGATATT 657  
 Db 598 AATGTTCATCATTTCCAGGTAAAGCTTTAGCTGAAATTAATAAATAATTTATGCTGACAATGAA 657  
 QY 658 GAGACCGTTGAGTATTTTCTCAACAGCAAAATCTTTGTCAGAGTGAACACATTTCT 717  
 Db 658 GAAGACATTTGATATCTCTTTGCTTCAACCAAGTTTATTTTAAAGTTGGAAATGTGAAC 717  
 QY 718 TTTTATACACCGCTCTTAGAAGGAAATTTATCCGATACAGACCGTTTATTAATGACAGAA 777  
 Db 718 TTTTATTTCTCAATTTTAGAAGGACATTTCTGATACACACGTTTATTTCTGAAAAC 777  
 QY 778 TTTGAGACGAGGTTGTTTTCATACCCCAATCCCTTCGCCACGCTATGGAACGTCCTTC 837  
 Db 778 TATGAAATTAATAAGTATAGACAAATGCGGAGTTTATCATCGATTTGATGCTGCACTCT 837  
 QY 838 TTGATTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTTGAGATTTACTCAAAATCATATT 897  
 Db 838 TTATTTAGCAGCGAAGGTTGTAATAACCGTTATTAATTTAAAGTACAGGTGATGACGTTGTT 897  
 QY 898 TCAGTCTATGTTAACTACCTCAGCTGAGTTGTTAAGGTAAACAGGAGTTTATAGATTTGTTAGT 957  
 Db 898 GAATATCTTCTACATCACCAGAAATTTGTTACTGTAAAAGAAAGTTGATGCAACGAT 957







DB 1078 AACCAAAAGGTGACGA 1094

RESULT 35

ACA20030

ID ACA20030 standard; DNA; 1134 BP.

XX AC ACA20030;

XX DT 19-JUN-2003 (first entry)

XX DE Prokaryotic essential gene #1687.

XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;

XX OS drug design; gene.

XX PN Staphylococcus aureus.

XX PD WO200277183-A2.

XX PF 03-OCT-2002.

XX PR 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall L, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI: 2003-029926/02.

XX DR P-PSDB; ABU16160.

XX PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 14; SEQ ID NO 7900; 1766pp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in

CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1134 BP; 411 A; 178 C; 194 G; 351 T; 0 U; 0 Other;  
Query Match 20.5%; Score 233; DB 8; Length 1134;  
Best Local Similarity 52.1%; Pred. No. 3.9e-40;  
Matches 571; Conservative 0; Mismatches 520; Indels 6; Gaps 2;  
QY 1 ATGATTCAATTTTCAATTAATCGCACATTATTTATTTATTCATGCTTTAAATACAACTAACCGT 60  
DB 1 ATGATGGAAATTCACATTTTAAAGAGATATTTTATACACATTTAAATACACATTTAAA 60  
QY 61 GCTATTAGCACTAAATAATGCCATTCCTATTTCTTTTCATCAATAATAAATGGAAGTCATCTT 120  
DB 61 GCTATTTCCCAAGAACCAACATTAACCTATATTTAACTGGTATCAAAATCGATCGGAAGAA 120  
QY 121 ACAGAGTAACCTTTTAACAGGGGTCTAACGGTCAATATCAATTTGAANAACATTTCTCTGTA 180  
DB 121 CATGAAGTTTATATACTAACTGGTTTCAGACTCTGAAATTTCAATAGAATCACTATCTAAA 180  
QY 181 AGT---AATGAAATGCTGGTTTGCTAATTAACCTCCAGGAGCTATTTTATTTAGAGCT 237  
DB 181 ACTGTAGATGGCGAGATATTTGTCATATTTTTCAGAAACAGGCTCAGTAGTACTTCTTGA 240  
QY 238 AGTTTTTTTATTAATATTTTCAAGTTTGGCAGATATTTAGTATTAATTTTAAAGAAATTT 297  
DB 241 CGATTCTTTGTTGATATTTATAAAAAAATTAACCTGGTAAAGATGTTTAAATTTATCTACAAAT 300  
QY 298 GAACAAACACCAAGTTGTTTAAACCAAGTGGTAAATCAGAGATTTACCTTAAAGAAAAGAT 357  
DB 301 GAACAAATTCACAGACATTAATTAACATCAGGTCAATCTGAAATTTTAAATTTAGTGGCTTAGAT 360  
QY 358 GTTGACCAAGTATCCTCGTCTACAGAGATATCAACAGAAATTCCTTTGATTTTAAAAACA 417  
DB 361 CCAGATCAATATCCTTTTATTTACCTCAAGTTTCTAGAGATGAGCAATTCATTTGTCGTA 420  
QY 418 AAATTTATTTGAAGTCTATTTATTTGCTGAAACAGCTTTTGGCAGCAGTTTACAAGAAAGTCT 477  
DB 421 AAAGTGCTTAAAAACGTATTTGCAACAAACGAATTTTGGCAGTGTCCACCTCCAGAAACACGC 480  
QY 478 CCTATTTAACAGGAGTTTCATATTTGTTAAGTATCATTAAGATTTTAAAGCAGTAGCG 537  
DB 481 CCAGTACTAACTGGTGGTGAATGCTGCTTATA---CAAGAAATGAAATTAATATGCAACAGC 537  
QY 538 ACTGACTCTCATCGTATGAGCCAAAGTTTAACTCACTTTTGGCAATACCTTTCAGCAGATTTG 597  
DB 538 ACTGACTCACACCGCTTGGCTGTAAAGAGTTGCAAGTTTGAAGATGTTTCTGAAACAAA 597  
QY 598 ATGGTAGTTTCTCCAAAGTAAATCTTTTGAGAGAAATTTTTCAGCAGTATTTTACAGATGATTT 657  
DB 598 AATGTCATCAATCCAGGTAAAGCTTTAGCTGAATTTAAATTTAAATTTATGCTGCAATGAA 657  
QY 658 GAGACCGTTGAGTATTTTCTCACCAGCCAAATCTTGTTCAGAGTGAACACATTTTCT 717  
DB 658 GAAGACATGATATCTCTTTGCTTCAACCAAGTTTATTTTAAAGTTGGAATGTGAAC 717  
QY 718 TTTTATACACGCTCTTAGAAGGAAATTTATCCGATACAGACCGTTTATTAATGACAGAA 777  
DB 718 TTTATTTCTCATTTATTAAGAGGACATTTATCTGTATACACACGTTTATTCCTGAAAC 777  
QY 778 TTTGAGACGGAGTTGTTTTCATACCAATCCCTTCGCCAGCTTATGGAACTGCGCTTC 837  
DB 778 TATGAAATTTAAATTAAGTATAGACAAATGGGAGTTTTATCATGCGATGATCGTGCCTCT 837  
QY 838 TTGATTTCTAATGCTACTCAAAATGGTACTGTTAGCTTTGAGATTTTACTCAAAATCATATT 897  
DB 838 TTAATTAGCACCGTGAAGGGTGAATTAACCGTTATTTAAATTTAAGTACAGGTGATGACGTTGTT 897  
QY 898 TCAGCTCATGTTTAACCTCACTGAGGTGCTGTAAGGTAAACGAGATTTAGATATTGTTAGT 957  
DB 898 GAAATTTCTTCTACATCACCAAGAAATTTGTTACTGTAAAAGAGAGTTGATGATCAACGAT 957

QY	958	CAGCTGGTAGTGAA	TTAACTATCACTGCTTCAATCCAACTTACCTTATTGAGTCTTTAAAA	1017		
Db	958	GTTGAAGGTGGTAG	CGCTGAAATTTTCATTCACTTAATATATGATGATGCTTTAAAA	1017		
QY	1018	GCTATTAAAGTCAAC	ACAGTAAATAATTCATTCTTATCACCAGTTTCGACCAATTCACCCCTA	1077		
Db	1018	GCATCGATAATGATG	AGGTTGAAGTTGAATTCCTTCGGTCAATGMAACCAATTTATTCTA	1077		
QY	1078	ACACGAGGCGATGAGGA	1094			
Db	1078	AAACCAAAAGGTGACGA	1094			
RESULT 36						
AAD62920						
ID	AAD62920 standard; DNA; 1134 BP.					
XX						
AC	AAD62920;					
XX						
DT	12-FEB-2004 (first entry)					
XX						
DE	Staphylococcus aureus dnaN gene beta subunit DNA.					
XX						
KW	Polymerase III enzyme; dnaE; dnaX; dnaB; Gram positive bacteria;					
KW	drug discovery; antibiotic activity; gene; ds.					
XX						
OS	Staphylococcus aureus.					
XX						
PH	Key	Location/Qualifiers				
FT	CDS	1..1134				
FT		/*tag= a				
FT		/product= "Staphylococcus aureus dnaN beta subunit"				
XX						
PN	US2003129633-A1.					
XX						
PD	10-JUL-2003.					
XX						
PF	28-OCT-2002; 2002US-00282287.					
XX						
PR	13-FEB-1998; 98US-0074572P.					
PR	22-JUL-1998; 98US-0093727P.					
PR	22-JAN-1999; 99US-00235245.					
XX						
PA	(ODON/) O'DONNELL M E.					
PA	(ZHAN/) ZHANG D.					
PA	(WHIP/) WHIPPLE R.					
XX						
PI	O'donnell ME, Zhang D, Whipple R;					
XX						
DR	WPI; 2003-829557/77.					
DR	P-PSDB; ABW01648.					
XX						
PT	New DNA replication proteins (i.e. subunits of the Staphylococcus aureus					
PT	DNA polymerase III enzyme) and genes, useful in drug discovery to screen					
PT	large libraries of chemicals for identification of compounds with					
PT	antibiotic activity.					
XX						
PS	Disclosure; Page 17-18; 69pp; English.					
XX						
CC	The invention relates to an isolated polypeptide, which comprises at					
CC	least one functionally active subunit of a Staphylococcus aureus DNA					
CC	polymerase III enzyme. The subunit comprises a 573 residue dnaE amino					
CC	acid sequence, a 566 residue dnaX amino acid sequence and/or a 457					
CC	residue dnaB amino acid sequence. The proteins and nucleic acids					
CC	replicate the chromosome of Gram positive bacteria and are useful in drug					
CC	discovery to screen large libraries of chemicals for identification of					
CC	compounds with antibiotic activity. The present sequence is S. aureus					
CC	dnaN gene beta subunit DNA					
XX						
SQ	Sequence 1134 BP; 411 A; 176 C; 194 G; 353 T; 0 U; 0 Other;					
Query Match 20.5%; Score 233; DB 10; Length 1134;						
Best Local Similarity 52.1%; Pred. No. 3.9e-40;						

	Matches	571;	Conservative	0;	Mismatches	520;	Indels	6;	Gaps	2
Qy	1	ATGATTCAATTTT	CAATTAATCGCA	CAATTTATTTATTT	CATGCTTTAAATCA	CACTAAACGT	60			
Db	1	ATGATGGAATTT	CACATTTAAAGAGAT	TATTTATTACCA	CAATTAATGACACAT	TAAAAA	60			
Qy	61	GCTATTAGCACT	AAAAATGCCATTC	TCTTCTTTATCA	TAAATAAAATGAAGCT	CACTTCT	120			
Db	61	GCTATTTCACCA	GAACAACATTTAC	CTATATTAAC	CTGGTATCAAAATCGATCGGAAGAA	120				
Qy	121	ACAGGAGTAGT	TAACACGGCTCTA	ACGGTCAAAATCA	ATTGTAAGAAACAC	TATTCCTGTGA	180			
Db	121	CATGAAGTTAT	TATTAACCTGGT	TCAGACTCTGAAAT	TTCAATAGAAATCACT	TATTCCTAAA	180			
Qy	181	AGT---	AATGAAAAATGCT	GGTTGCTAA	TACTCTCCAGGAGCT	TATTTATTAGAACT	237			
Db	181	ACTGTAGATGG	CGAAGATATCT	CAATATTT	CAGAAACAGGCTCAGTAGT	ACTTCTCGGA	240			
Qy	238	AGTTTTTTTT	TATTAATATATTT	CAAGTTTGC	CAGATATTAGTA	TAAATGTTTAAAGAAAT	297			
Db	241	CGATTCCTTT	GTGTATTTATA	AAAAAAATTA	CTGGTAAAGATGTTAAAT	TATCTACAAAT	300			
Qy	298	GAACAAACCA	AGTTGTTTAA	CCAGTCTGTAAT	TCAGAGATTCCTTAA	AAAGAAAGAT	357			
Db	301	GAACAATTC	CAACATTAAT	TACATCAGCT	CAITCTGAA	TTAATTTACTACAAAT	360			
Qy	358	GTGTACCGAT	ATCCTCGTCTA	CAAGAAGTATCA	ACAGAAATCCTTTG	ATTTAAAAACA	417			
Db	361	CCAGATCAAT	ATCCTTTATTA	CCCTCAAGTTT	CTAGAGATGACG	CAATTCATTTCTCGGTA	420			
Qy	418	AAATTTATTT	GAAAGTCTAT	TATTTGCTG	AAACAGCTTTTTC	CGCCAGTTTAC	477			
Db	421	AAAGTACTT	AAAAACGTGAT	TGCAACAGAA	TTTGGCAGTTTCC	ACTCAGAAACACGC	480			
Qy	478	CCTATTTTAA	CAGAGTTTCA	TATGTATTA	AGTAATCATAA	AGATTTTAAAGCAGTAGCG	537			
Db	481	COAGTACTAA	CTGGTGTGA	ACTGGCTTTATA	---CAAGAAAATGA	ATTAATATGACACAGC	537			
Qy	538	ACTGACTCT	CATCTGTATG	AGCCAAAGTTT	TAATCACATTTG	CACATATCTTACAGAGATTTG	597			
Db	538	ACTGNATTC	ACACCCGCTT	GGCTGTAG	AAAGTTG	CAGTTAGAAATGTTTCTG	597			
Qy	598	ATGGTAGTCT	TTCCAAGTAAAT	CTTTGAGAGAA	TTTTTTCAGCAGT	TATTTACAGATGATATT	657			
Db	598	AATGTCTCAT	TTCCAGGTAA	GGCTTTTAGCTGA	ATTAATAAATATATG	CTGACATGAA	657			
Qy	658	GAGACCGT	TGAGGTATTTT	CTCACCAGCCAA	TCTTGTTCAGA	AGTGAACACATTTCT	717			
Db	658	GAAGACAT	TGATACTCTT	CTTGTCTCAAC	CAAGTTTATTTAA	AGTTGGAATGTGAAC	717			
Qy	718	TTTTATAC	CGCCTCTTAG	AGGAAATTA	TCCCGATACAG	CCGTTTATTTAATGACAGAA	777			
Db	718	TTTTATTTCT	CGAATATTAGA	AGGACATATTC	CTGATACAA	CAGTTTATTCCTGAAAC	777			
Qy	778	TTTGAGACG	GGTGTGTTT	CTCATACCC	AAATCCCTTCG	CCACGCTATGAA	837			
Db	778	TATGAAAT	TAAATTAAGAT	ATAGACAAT	TGGGAGTTTATCAT	ATCGCATTCATCGTCTCT	837			
Qy	838	TTGATTTCT	TAATGCTACT	CAAAAATGG	TATGTTAAGCTTTG	AGATTTACAAATCATATT	897			
Db	838	TTATTTAG	CAGTCAAGGTG	TATTAACGTTAT	TAAATTAAGTAC	AGGTGATGACGTTGTT	897			
Qy	898	TCAGCTCAT	GTAACTCA	CTGAGGTTGG	TAAAGTTAA	ACGAGATTTAGATATGTTAGT	957			
Db	898	GAATTAATCT	TTCATCATCA	CCAGAAAT	TGGTACTGT	TAAAGAAAGAAAGTTGATG	957			
Qy	958	CAGTCTGGT	AGTGATTTAA	CTACTCAGT	TTCAATC	CAATTTACCTTTATGAGTCTTTAAA	1017			
Db	958	GTTGAAGG	TGGTAGCCTG	AAATTTCA	TTCACTCTAA	TATATGATGATGCTTTAAA	1017			
Qy	1018	GCTATTTAA	AGTGAAACAG	TAAAAAAT	TCATTTCTTAT	CACAGTTTCGACCACTTAC	1077			
Db	1018	GCAATCGA	TAAATGATGAG	GTGAAGTTGA	ATTTCTTCGGT	TACAAATGAAACCAATTTCTA	1077			



```

Db      1078 CCAAAAGGTGACGA 1091
|||||
RESULT 38
ABQ70950
ID      ABQ70950 standard; DNA; 4736 BP.
XX
AC      ABQ70950;
XX
DT      29-AUG-2003 (revised)
DT      29-AUG-2002 (first entry)
XX
DE      Listeria monocytogenes 4b contig DNA sequence #892.
XX
KW      Antibacterial; Listeria; food contamination; mutational analysis;
KW      infection; ds.
XX
OS      Listeria monocytogenes ATCC 19115.
XX
PN      WO200228891-A2.
XX
PD      11-APR-2002.
XX
PF      04-OCT-2001; 2001WO-FR003061.
XX
PR      04-OCT-2000; 2000FR-00012697.
XX
PA      (INSP ) INST PASTEUR.
PA      (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI      Kunst F, Glaser P;
XX
DR      WPI; 2002-332479/37.
XX
PT      New genomic sequences from Listeria species, useful for detection,
PT      treatment and prevention of infection, also related polypeptides,
PT      antibodies and modulators.
XX
PS      Claim 14; SEQ ID NO 3763; 180pp; French.
XX
CC      The present invention relates to nucleic acid sequences (ABQ67196-
CC      ABQ71212) from Listeria sp. The sequences are useful as probes and
CC      primers for identification and/or detection of Listeria (e.g. as
CC      contaminants in foods, or mutational analysis) and for analysis of gene
CC      expression. Proteins encoded by the nucleic acid sequences can be used to
CC      screen for compounds that modulate gene expression, replication and
CC      pathogenicity of Listeria (potential therapeutic agents), also for
CC      treating infections by Listeria, and are useful as immunogens in anti-
CC      Listeria vaccines. Note: The sequence data for this patent did not form
CC      part of the printed specification, but was obtained in electronic format
CC      directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. (Updated
CC      on 29-AUG-2003 to standardise OS field)
XX
SQ      Sequence 4736 BP; 1521 A; 800 C; 938 G; 1477 T; 0 U; 0 Other;

Query Match      17.7%; Score 200.4; DB 6; Length 4736;
Best Local Similarity 51.5%; Pred. No. 4.5e-33;
Matches 565; Conservative 0; Mismatches 521; Indels 12; Gaps 4;

QY      2 TCATTCAATTTTCAATTAATCGCAATTTATTTCATGCTTTTAAATACAACTAAACGTG 61
Db      1218 TCATGAATTTTGTATTGAGCGGTGATCGTCTGTCGAAGCAGTCAATGAAGTACTCGTG 1277
QY      62 CTAATTAGCACTAAATAATCCCAATTCCTATTCTTTTCATCAATAAATAATGAAGTCACTCTA 121
Db      1278 CCATCTCTGCAAGACACAGCAATTCATTTTAAACGGGATAAATAATAGTGTAAATGATG 1337
QY      122 CAGGAGTAACCTTTAACAGGGGTCTAACGGGTCAAAATATCAATTAATGAAACACTATTCCTGTAA 181
Db      1338 AAGGTGTAACACTAACTAGGTAGTATTCCGATATTTCAATCGAAGCAATTTATTCATTAA 1397
QY      182 GT---AATGAAATGCTGGTTTGTCTAATTACCTCTCCAGGAGCTATTTTATTAGAACTA 238

```

```

Db      1398 TTGAAAATCATGAGTAATTTAGAGTGGAGAGTTTTGGTGAATTTGTACTTCAATCAA 1457
QY      239 GTTTTTTTTAAATTAATATTTCAGAGTTTGGCCAGATATTAGTATAAATGTTTAAAGAAATG 298
Db      1458 AATACCTTTGGCGATATTGTTCTGTTTACCAGAGAAATATGTAAGAAATGAAGTGACTT 1517
QY      299 AACACACCAACCAAGTTGTTTTAACCCAGTGGTAAATCAGAGATTACCTTTAAAGAAAAGATG 358
Db      1518 CTAACCTACCAACCAACCAATAGTTCTGGCCAGCAATCTTTTACATTAATGGCTTAGATC 1577
QY      359 TTGACCGATATCTCTGCTACAGAAAGTATCAACAGAAAATCCCTTTGATTTTAAACAA 418
Db      1578 CAATGGAATATCTTAAATTTACCTGAAGTAAACAGACGGAACAAATTAATTTCAATTA 1637
QY      419 AATTATTGAAGTCTATTATTGCTGNAACAGCTTTTGCAGCCAGTTTCAAGAAAGTCGTG 478
Db      1638 ATGTACTTAAATAATTTGTTAGACAAATCTGTTTGTCTGCTGCGATTGAAGTTCGTG 1697
QY      479 CTATTTTAAACAGGAGTTTCATATTGTTAATTAAGTAATCAATAAGATTTTAAAGCAGTAGCGA 538
Db      1698 CAGTACTTACTGGTGTAACTGGATTTCAAGAAAATTA---CTAAGCGCAGTTGC 1754
QY      539 CTGACTCTCATCTATGAGCCCAAGCTTTAATCACTTTTGGACA---ATACTTTCAGCAGATT 595
Db      1755 CCGATAGTCACTGCTAGCTTTTACGTGAATACCTCTTGAACAGACATTTGATGAAGAAT 1814
QY      596 TGATGGTAGTTCTTCCAAAGTAAATCTTTGAGAGAAATTTTCAGCAGATTTTACAGATGATA 655
Db      1815 ACAACATTGTTTCTCTGGAATAAGTTTATCTGAATTAATAAATCTTTTATGATGATGCAA 1874
QY      656 TTGAGACGGTTGAGGTATTTTCTCACAAGCCAAATCTTTGTTTCAAGAGTGAACACATTT 715
Db      1875 GCGAATCTATTGAAATGACCCCTTGCACAAACCAAAATCTTTTAAATTAAGATTTAT 1934
QY      716 CTTTATTATACCGCTCTTAGAAGGAAATTTATCCGATACAGACCGTTTATTAATGACAG 775
Db      1935 TATTTTATTCTGTTTACTTGAAGGTAGTTATCCAGATACATCTCGATTAAATTTCCAAC 1994
QY      776 AATTGAGACGGAGTTGTTTTCATACCCAAATCCCTTCGCCACGCTATGGAAACGTGCCT 835
Db      1995 ATACTAAATCAGAAATTAGTCAATTAATCCAAAGCAATTTTACAGCAATGACCGTGCCT 2054
QY      836 TCTTGATTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTGAGATTAATCTCAAAAT---C 892
Db      2055 CCTACTTGTCTCCGAAATTCGTAATACGTTTATTAATTAATGACGCTAGAAAATGGCC 2114
QY      893 ATATTTCACTCATGTTTAACTCACTCCTGAGTTTGGTAAAGTAAACGAGGATTTAGATATTG 952
Db      2115 AAGTAGAAGTATCATCCAAATCTCCAGAGTTGGGAAATGTTTCTGAAAATGCTTCAGCC 2174
QY      953 TTAGTCAGTCTGCTAGTGTATTAATCATCAGCTTCAATCCAACTTACCTTATTTAGTCTT 1012
Db      2175 AAGTTTTTACTGGCGAAGAGATCAAAATATCTTTTAAACGGTAAATATACATGATGGATGCC 2234
QY      1013 TAAAGCTATTAAAGTGAACAGATAAAATTCATTTCTTATCACCAGTTTCAGCCATTC 1072
Db      2235 TAGGTGCTTTTGAAGTGTATGATATTCAAATTTCTTTCCTCGGTACAAATGACCAATTCG 2294
QY      1073 CCCTAACACCGCGCATG 1090
Db      2295 TACTTCGACCAAAAGATG 2312

RESULT 39
ABQ67196.2/c
Continuation (3 of 7) of ABQ67196 from base 200001 (Listeria innocua contig DNA sequence
WP Sequence split into 7 fragments LOCUS ABQ67196 Accession ABQ67196
WP Fragment Name Begin End
WP ABQ67196.0 1 110000
WP ABQ67196.1 100001 210000
WP ABQ67196.2 200001 310000
WP ABQ67196.3 300001 410000

```

WP	ABQ67196_4	400001	510000
WP	ABQ67196_5	500001	610000
WP	ABQ67196_6	600001	684707
Query Match			
Best Local Similarity 17.5%; Score 198.8; DB 6; Length 110000;			
Matches 564; Conservative 0; Mismatches 522; Indels 12; Gaps 4;			
Qy	2	TGATTCAAATTTTCAATTAATTCGCACATAATTAATTATTCATGCTTTAAATACAACCTAAACGTG	61
Db	67740	TCATGAATTTTGTTATTAGCGTGATCGCTTGTGCCAAGCAGTCAATGAAGTTACTCTGTG	67681
Qy	62	CTATTAGCACTAAAATGCCATTCCTATTCTTTTCATCAATAAAAAATTGAAGTCACCTTCTA	121
Db	67680	CCATCTCTGCAAGAACCAACGATTTCCAATTTTAAACGGGATAAAAAATAGTCGTAATATGATG	67621
Qy	122	CAGGAGTAACTTTFAACAGAGTCTAACCGGTCAAAATATCAATATGAAACACACTATTCTCTGTAA	181
Db	67620	AAGGTGTACTTTTAAACCGGTAGGGATCTGTGATTTTCCATCGAAGCATTTATTCATTTAA	67561
Qy	182	---GTAATGAAAAATGCTGGTTTGTCTAAATACCTCTCCAGGAGCTATTTTTATTAGAAGCTA	238
Db	67560	TCGAAAATGATGAAGTAAATTGAGAGTTGAAAGTTTTCGTGGTATCGTCTTCAATCCA	67501
Qy	239	GTTTTTTTATTAATATATTTCAGTTTCGCAGATATATAGTATATAATGTTTAAAGAAATGT	298
Db	67500	AGTACTTCGGTGTATATTGTTTCGTGTTTACCCTGAAGAAAATGTCGAAAATGGAAGTTACTA	67441
Qy	299	AACAACACCAAGTTGTTTAAACGAGTGTAATACAGAGATTACCTTTAAAGGAAAAGATG	358
Db	67440	CAAAATTATCAAACTAACATTAGTTCTGGCCAGCTTCTCTTTACACTAACCGCTTAGATC	67381
Qy	359	TTGACCAGTATCCTCGTCTACAAGAGTATCAACGAAAAATCCTTTGATTTTTAAAAACAA	418
Db	67380	CAATGGAATACCCTAAATTACCAGAAGTTTACTGATGTAABAATATTAATTTCTCTATTA	67321
Qy	419	AATATTGAAGTCTATTATTGCTGAAACAGCTTTTTCGAGCCAGTTTCAAGAAAGTCGTG	478
Db	67320	ATGCTCTTAAAAACATTAATTAGACAAACTGTTTTTTCGTCTGCTATCGAAGTTTCGCC	67261
Qy	479	CTATTTTACAGAGTTTCATATTGTTATTAAAGTAATCATAAAGATTTTTAAAGCAGTAGCGA	538
Db	67260	CAGTGCTTACAGGTGTAACCTGGATTAT---TAAAGATAACAAACTTAGCGCAGTTTGCAA	67204
Qy	539	CTGACTCTCATCTATGAGCCAACTGTTAATCACCTTTTGA---CAATACTTCAGCAGATT	595
Db	67203	CCGATAGTCATCGCTAGCATTTACGTGAATTTCCACTTGAAACAATATTGACGAAGAT	67144
Qy	596	TGATGGTAGTTCCTCAAGTAAATCTTTGAGAGAAATTTTCAGCAGTATTTACAGATGATA	655
Db	67143	ACAACATTTGTTATTCAGGAAAAAGCTTAGCTGAACATAAAATTTTAGACGATGCAA	67084
Qy	656	TTGAGACCGTTGAGGTATTTTCTCACCAAGCCAAATCTTGTTCAAGAGTGACACATTT	715
Db	67083	GTGAATCCATTTGAAATGACATCGCAAAATAATCAAAATCTATTTAATTTAAAGATTTAT	67024
Qy	716	CTTTTTTATACACGCCCTCTTAGAAGGAAATTTATCCCAGATACAGACCGTTTATTAAATGACAG	775
Db	67023	TATTTTATCTCGTTTACTGGAAGGTAGTTTACCAGATACATCACGCTTAATTTCCAATG	66964
Qy	776	AATTTGAGACGGAGTTGTTTTTCAATACCCAATCCCTTCGCCACGCTATGGAACGTGCT	835
Db	66963	ATACTAAATCAGAACTTGTCATTAAATTCAGAGCATTTTTTACAAGCATTTGACCGTCTT	66904
Qy	836	TCCTTGATTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTGAGATTACTCAAAAT---	892
Db	66903	CCCTACTTGTCTCGGAAAAACCGTAATAACGTTTATTAAATTAATGATGCCCTTTGAAAAATG	66844
Qy	893	ATATTTTCAGCTCATGTTTAACCTCACCTGAGGTGTTAAGTAAAGTAAACGAGGATTTAGATAT	952
Db	66843	AAGTTGAAGTATCCTCTAACTCTCTCGAAGTTGGGAATGTTTCGGGAAAATGCTTTAGTTC	66784
Qy	953	TTAGTCAGTCTGGTAGTGAATTTAACTATCAGCTTCAATCCAACTTACCTTATTGGAGTCTT	1012

DR WPI; 2002-332479/37.  
XX New genomic sequences from *Listeria* species, useful for detection,  
PT treatment and prevention of infection, also related polypeptides,  
PT antibodies and modulators.  
XX  
PS Claim 5; SEQ ID NO 2058; 180pp; French.  
XX  
CC The present invention relates to nucleic acid sequences (AB067188-  
CC AB071212) from *Listeria* sp. The sequences are useful as probes and  
CC primers for identification and/or detection of *Listeria* (e.g. as  
CC contaminants in foods, or mutational analysis) and for analysis of gene  
CC expression. Proteins encoded by the nucleic acid sequences can be used to  
CC screen for compounds that modulate gene expression, replication and  
CC pathogenicity of *Listeria* (potential therapeutic agents), also for  
CC treating infections by *Listeria*, and are useful as immunogens in anti-  
CC *Listeria* vaccines. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 3011208 BP; 941651A; 568176C; 559189G; 942192T; 0U; 0Other;  
Query Match 17.5%; Score 198.8; DB 6; Length 110000;  
Best Local Similarity 51.4%; Pred. No. 1.7e-32;  
Matches 564; Conservative 0; Mismatches 522; Indels 12; Gaps 4;  
  
QY 2 TGATTCATTTTCAATTAATCGCAGTATTTATTCATGCTTTTAAATACAACTAAACGTTG 61  
DB 1866 TCATGAATTTGTTATTTAGCGTGTGCTTGTCCAGCAGTCAATGAAGTTACTCGTG 1925  
  
QY 62 CTATTAGCAGTAAATGCAATTCCTATTTCTTCATCAATTAATTAAGTCACTTTCTA 121  
DB 1926 CCACTCTCGAAGACACACGATTCCTCAATTTTAAACGGGATTAATAATAGTCGTAATGATG 1985  
  
QY 122 CAGAGTAACTTTAAACAGGCTCTAACGGTCAATATCAATTCGAAACACATTTCTCTGTAA 181  
DB 1986 AAGGTGTACTTTAAACGGTAGCGATTCTGATTTTCCATCGAAGATTTATTCATTTAA 2045  
  
QY 182 ---GTAATGAATGCTGGTTGGTAACTTACCTCTCCAGGAGCTATTTTATAGAGCTA 238  
DB 2046 TCGAAATGATGAAGTAAATTTAGAGTTGAAAGTTTGGTGGTATCGTTCTTCAATCCA 2105  
  
QY 239 GTTTTTTAAATTAATTTATTTCAAGTTTGCAGATATTAGTATTAATGTTAAAGNAATTG 298  
DB 2106 AGTACTTCGGTGATATTGTTCTGTTTACCTGAGAAATGTCGAAATTTGAAGTTACTA 2165  
  
QY 299 AACACACCAAGTTGTTTAAACAGTGGTAAATCAGAGATTACCTTTAAAGGAAAGATG 358  
DB 2166 CAAATTTACAACTAACATTAAGTTCTGGCCAGCTTCTTTACACTAAACGGCTTAGATC 2225  
  
QY 359 TTGACCAGTATCTCGTTACAGAGATGATCAACAGAAATCCTTTGATTTTAAACAA 418  
DB 2226 CAATGGAATACCTAAATTTACAGAAATTTCTGATGGTAAATAATTTAAATTCCTATTA 2285  
  
QY 419 AATTATTGAAGTCTAATTTGCTGAACAGCTTTTGACCCAGTTTCAAGAAAGTCGTC 478  
DB 2286 ATGTCCTTTAAACCAATTTATAGACAACTGTTTGTGCTTCTGCTATCGAAGTTGCCC 2345  
  
QY 479 CTATTTTAAACAGGCTTCATATTGTTAATTAATCATAAAGATTTTAAAGCAGTAGCGA 538  
DB 2346 CAGTCTTTACAGGTGTTAACTGGATTAT---TAAAGATAACAAACTTAGCGCAGTTGCAA 2402  
  
QY 539 CTGACTCTCATCGTATGAGCCAAAGTTTAAATCACTTTTGA---CAATACTTCAGCAGATT 595  
DB 2403 CCGATAGTATCGTCTAGCATTACGTGAATTTCCATTTGAACAAATTTAGCAGAAAT 2462  
  
QY 596 TGATGGTAGTTCTTCCAGTAATTTCTTTGAGAAATTTTTCAGAGTATTTACAGATGATA 655  
DB 2463 ACAACATTTGTTTCCAGGAAAGAGCTTAGCTGAACCTAAATAAAATTTTAGACGATGCAA 2522  
  
QY 656 TTGAGACCGTTGAGGTATTTTCTCACCAGCCAAATCTTTGTCAGAGTGAACACATTT 715  
DB 2523 GTGAATCCATTTGAATGACACTCGCAATTAATCAATTCCTATTTTAAATTAAGATTTAT 2582

QY 716 CTTTTTATACAGCCTCTTTAGAGGAAATTTATCCGATACAGACGCTTTTATTATGACAG 775  
DB 2583 TATTTTATCTGTTTACTTGAAGGTAGTTACCCAGATACATCAGCTTATTTCCACTG 2642  
  
QY 776 AATTTGAGACGGAGGTTGTTTCAATACCCATCCCTTCGCCAGCTATGGAAGTGCCT 835  
DB 2643 ATACTAAATCAGAACTTGTCTATTAATCCAGAGCATTTTACAAAGCGATTGACCGTCTT 2702  
  
QY 836 TCTTCATTTTAAATGCTACTCAAAATGCTACTGTTAAGCTTGAGTTACTCAAAAT---C 892  
DB 2703 CCTACTTCTCGCGAAACCGTAAATACGTTATTAATTAATGACCCCTTGAATGGAC 2762  
  
QY 893 ATATTTAGCTCATGTTTAACTCACCTGAGGTGGTAAAGGTAACGAGGATTTAGATATTG 952  
DB 2763 AAGTTGAAGTATCTCTTAACCTCTCTGAAGTTGGGAATGTTTCGGAATAATGCTTTAGTC 2822  
  
QY 953 TTAGTCAGTCTGGTAGTGATTTAACTATCATGCTTCAATCCAACTTACCTTATTGAGTCTT 1012  
DB 2823 AAAGCTTTACTGGTGAAGAAATTAATAATATCCCTTTAAACGGTAAATATCATGATGCAT 2882  
  
QY 1013 TAAAGCTATTAAAGTGAAACAGTAAATAATTTCAATTTCTTATCACCAAGTTTCGACATTCA 1072  
DB 2883 TCGCGCATTTGAAGGCGATGATATTCAAAATCTCTTCTCTGGCACATGAGACCTTTTCG 2942  
  
QY 1073 CCCTAACACACGAGCGATG 1090  
DB 2943 TACTTCGACCAAAAGATG 2960

Search completed: January 31, 2005, 10:59:43  
Job time : 447 secs

GenCore version 5.1.6  
Copyright (C) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2005, 10:52:07 ; Search time 3366 Seconds  
(without alignments)  
15931.824 Million cell updates/sec

Title: US-10-048-071-27  
Perfect score: 1134  
Sequence: 1 atgattcaatttcaattaa.....ttacaccagtagcaaac 1134

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_btg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sv.\*  
13: gb\_un.\*  
14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1134	100.0	1134	1	AF280765 Streptoco
2	1126	99.3	1134	6	CQ648158 Sequence
3	1126	99.3	12551	1	AE006472 Streptoco
4	1118	98.6	110000	1	CP000003 Streptoco
5	1116.4	98.4	14943	1	AE009953 Streptoco
6	1114.8	98.3	51896	1	AE014136 Streptoco
7	1114.8	98.3	311600	1	AP005141 Streptoco
8	849	74.9	3200	6	AR157860 Sequence
9	849	74.9	3200	6	E17380 gDNA derive
10	690.8	60.9	14925	1	AE014853 Streptoco
11	641.2	56.5	1134	6	CQ648156 Sequence
12	641.2	56.5	1137	6	AX606582 Sequence
13	641.2	56.5	21262	1	AE014191 Streptoco
14	641.2	56.5	26144	6	AX602184 Sequence
15	641.2	56.5	60266	6	CQ655074 Sequence
16	641.2	56.5	143050	1	SAG766843 Streptoco
17	641.2	56.5	349980	6	CQ655073 Sequence
18	641.2	56.5	349980	6	AX954527 Sequence
19	608.2	53.6	1134	6	AX571491 Sequence

20	608.2	53.6	1137	6	AR480840
21	608.2	53.6	14418	1	AE007318 Streptoco
22	608.2	53.6	21338	6	CQ788929 Sequence
23	608.2	53.6	21338	6	AR218788 Sequence
24	608.2	53.6	21338	6	BD003700 Polynucle
25	608.2	53.6	232807	2	SPNEU1901 Streptoco
26	608.2	53.6	349980	6	AX571766 Sequence
27	603.4	53.2	6098	1	SPDNAARG Streptoco
28	603.4	53.2	14222	1	AE008385 Streptoco
29	569	50.2	2876	1	AF255728 Streptoco
30	408.2	36.0	10186	1	AE006239 Lactococc
31	408.2	36.0	10762	1	LLU76424 Lactococc
32	339.4	29.9	326434	1	AE016947 Enterococ
33	330.6	28.2	8001	6	BD194028 Enterococ
34	317	28.0	1155	6	AR346291 Sequence
35	252	22.2	343050	1	AL935252 Lactobaci
36	247.4	21.8	300787	1	AE016744 Staphyloc
37	242	21.3	300007	1	AE017200 Lactobaci
38	242	21.3	349980	6	AX926711 Sequence
39	236.2	20.8	1131	6	AX621126 Sequence
40	236.2	20.8	1134	6	AX554149 Sequence
41	236.2	20.8	1134	6	AX741532 Sequence
42	236.2	20.8	1134	6	AX741534 Sequence
43	236.2	20.8	2347	6	AR354035 Sequence
44	236.2	20.8	49336	1	AC025950 Staphyloc
45	236.2	20.8	50152	1	AC025948 Staphyloc

#### ALIGNMENTS

RESULT 1  
AF280765  
LOCUS  
DEFINITION Streptococcus pyogenes DNA polymerase III beta subunit (dnaN) gene, partial cds.  
ACCESSION AF280765  
VERSION AF280765.2 GI:16885214  
KEYWORDS Streptococcus pyogenes  
SOURCE Streptococcus pyogenes  
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.  
REFERENCE 1 (bases 1 to 1134)  
AUTHORS Bruck, I. and O'Donnell, M.  
TITLE The DNA Replication Machine of a Gram-Positive Organism  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1134)  
AUTHORS Bruck, I. and O'Donnell, M.  
TITLE Direct Submission  
JOURNAL Submitted (21-JUN-2000) DNA Replication, Rockefeller University/HHMI, 1230 York Ave, New York, NY 10021, USA  
REFERENCE 3 (bases 1 to 1134)  
AUTHORS Bruck, I. and O'Donnell, M.  
TITLE Direct Submission  
JOURNAL Submitted (09-NOV-2001) DNA Replication, Rockefeller University/HHMI, 1230 York Ave, New York, NY 10021, USA  
REMARK Sequence update by submitter  
COMMENT On Nov 9, 2001 this sequence version replaced gi:9789548.  
FEATURES  
Location/Qualifiers  
source  
1..1134  
/organism="Streptococcus pyogenes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1314"  
1..1134  
/gene="dnaN"  
1..1134  
/gene="dnaN"  
/transl\_table=1  
/product="DNA polymerase III beta subunit"  
/protein\_id="AAF98349.2"  
/db\_xref="GI:16885215"



/translation="MIQSFNRLFIHALNTTKRAISKNAIPILSSIKIEVSTGT  
LTSNGOISIENTIPVSNENAGLLITSPAILLEASFFINIISUPDISINVKIEQH  
QVILSGSEITLKGNDQVPRLEVENSTENPLIKTLKLSIIAETFAASLOESRP  
ILTVHIVLNSHDKFKAVIDSHRMSQRLITLNDTSADLMVVLFSKISREFAVFTDD  
IETVFFSPQILFRSEHISYFLLEGNVPTDRLLMTBFEVVENTSLRHHAME  
RAFLSNATQNGTVELEITQNHISAHVANSPEVKYNEDLDIVSQSGDLTISFNPTYL  
IESLKAIRSETVKIHFSPVRFTLTPGDEBESFQLITPVRTN"

ORIGIN

Query Match 100.0%; Score 1134; DB 1; Length 1134;  
Best Local Similarity 100.0%; Pred. No. 1.4e-198;  
Matches 1134; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATTCATTTTCAATTAATCGCACATTTATTTATTCATGCTTTTAAATCAACTAAACGT 60  
DB 1 ATGATTCATTTTCAATTAATCGCACATTTATTTATTCATGCTTTTAAATCAACTAAACGT 60

QY 61 GCTATTAGCACTAAATGCAATTCCTATTCTTTTCATCAATAAATAATGCAATCTTCT 120  
DB 61 GCTATTAGCACTAAATGCAATTCCTATTCTTTTCATCAATAAATAATGCAATCTTCT 120

QY 121 ACAGGAGTAACCTTAAACAGGCTTAACGGTCAATATCAATTCAGAACACTATTCCTGTA 180  
DB 121 ACAGGAGTAACCTTAAACAGGCTTAACGGTCAATATCAATTCAGAACACTATTCCTGTA 180

QY 181 AGTAATGAAATGCTGCTGTTGCTAAATCTCTCCAGGAGCTATTTTATTAAGAGTACT 240  
DB 181 AGTAATGAAATGCTGCTGTTGCTAAATCTCTCCAGGAGCTATTTTATTAAGAGTACT 240

QY 241 TTTTATTAATTAATTTCAAGTTTGGCAGATATTAGTATTAATTTAAAGAAATGAA 300  
DB 241 TTTTATTAATTAATTTCAAGTTTGGCAGATATTAGTATTAATTTAAAGAAATGAA 300

QY 301 CAACACCAAGTCTTTTAAACAGGCTTAATCAGAGTATACCTTTAAAGGAAAGATGT 360  
DB 301 CAACACCAAGTCTTTTAAACAGGCTTAATCAGAGTATACCTTTAAAGGAAAGATGT 360

QY 361 GACGAGTATCCTGCTTACAGAAAGTATCAACAGAAATCCTTTTGATTTTAAAAA CAAA 420  
DB 361 GACGAGTATCCTGCTTACAGAAAGTATCAACAGAAATCCTTTTGATTTTAAAAA CAAA 420

QY 421 TTATTGAAGTCTATTATTTGCTGAAACAGCTTTTGGCCAGCTTTTACAGAAAGTCTCT 480  
DB 421 TTATTGAAGTCTATTATTTGCTGAAACAGCTTTTGGCCAGCTTTTACAGAAAGTCTCT 480

QY 481 ATTTTAAACAGGAGTCTATTTGATTTAGTAATCAATAAGATTTTAAAGCAGTAGGACT 540  
DB 481 ATTTTAAACAGGAGTCTATTTGATTTAGTAATCAATAAGATTTTAAAGCAGTAGGACT 540

QY 541 GACTCTCATCGTATGAGCAACGTTTAAATCACTTTTGACAAATCTTTCAGCAGATTTGATG 600  
DB 541 GACTCTCATCGTATGAGCAACGTTTAAATCACTTTTGACAAATCTTTCAGCAGATTTGATG 600

QY 601 GTAGTCTTCCAGTAATCTTTGAGAGAAATTTTTCAGCAGTATTTTACAGATGATATGAG 660  
DB 601 GTAGTCTTCCAGTAATCTTTGAGAGAAATTTTTCAGCAGTATTTTACAGATGATATGAG 660

QY 661 ACCGTTGAGGATTTTCTCAGCAAGCCAAATCTTTGTTTCAAGAGTGAACATTTCTTTT 720  
DB 661 ACCGTTGAGGATTTTCTCAGCAAGCCAAATCTTTGTTTCAAGAGTGAACATTTCTTTT 720

QY 721 TATACAGCCTCTTAGAAGGAAATTTATCCCGATACAGACCGTTTATTAATGACAGAAATTT 780  
DB 721 TATACAGCCTCTTAGAAGGAAATTTATCCCGATACAGACCGTTTATTAATGACAGAAATTT 780

QY 781 GAGACGAGGATTTTCTCAATACCCATCCCTTCGCCAGCTATGGAACGCTCTCTTG 840  
DB 781 GAGACGAGGATTTTCTCAATACCCATCCCTTCGCCAGCTATGGAACGCTCTCTTG 840

QY 841 ATTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTCAGATTAATCAAAATCATATTTCA 900  
DB 841 ATTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTCAGATTAATCAAAATCATATTTCA 900

QY 901 GCTCATGTTAACTCACTGAGGTTGGTAAGTAAACGAGGATTTAGATATTTAGTCTAG 960  
DB 901 GCTCATGTTAACTCACTGAGGTTGGTAAGTAAACGAGGATTTAGATATTTAGTCTAG 960

QY 961 TCTGCTAGTAAATTAACATATCAGCTTCAATCCAACTTACCTTTTGTAGTCTTTAAAAGCT 1020  
DB 961 TCTGCTAGTAAATTAACATATCAGCTTCAATCCAACTTACCTTTTGTAGTCTTTAAAAGCT 1020

QY 1021 ATTTAAAGTGAACAGATAAATAATTCATTTCTTATCACCAGTTTCAGCAATTCACCTTAACA 1080  
DB 1021 ATTTAAAGTGAACAGATAAATAATTCATTTCTTATCACCAGTTTCAGCAATTCACCTTAACA 1080

QY 1081 CCAGCGATGAGGAAGAAAGTTTATCCAAATTAATTAACACAGTACGAAACAAAC 1134  
DB 1081 CCAGCGATGAGGAAGAAAGTTTATCCAAATTAATTAACACAGTACGAAACAAAC 1134

RESULT 2  
LOCUS CQ648158 1134 bp DNA linear PAT 02-FEB-2004  
DEFINITION Sequence 5115 from Patent WO0234771.  
ACCESSION CQ648158  
VERSION CQ648158.1 GI:41683805  
KEYWORDS Streptococcus pyogenes  
SOURCE Streptococcus pyogenes  
ORGANISM Streptococcus pyogenes  
REFERENCE Telford, J., Masignani, V., Margarit y Ros, I., Grandi, G., Fraser, C. and Tettelin, H.  
AUTHORS Nucleic acids and proteins from streptococcus groups a & b  
TITLE Patent: WO 0234771-A 5115 02-MAY-2002;  
JOURNAL Chiron S.p.A. (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)  
FEATURES  
Location/Qualifiers  
1..1134  
/organism="Streptococcus pyogenes"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1314"

ORIGIN

Query Match 99.3%; Score 1126; DB 6; Length 1134;  
Best Local Similarity 99.6%; Pred. No. 4.2e-197;  
Matches 1129; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATGATTCATTTTCAATTAATCGCACATTTATTTATTCATGCTTTTAAATCAACTAAACGT 60  
DB 1 ATGATTCATTTTCAATTAATCGCACATTTATTTATTCATGCTTTTAAATCAACTAAACGT 60

QY 61 GCTATTAGCACTAAATGCAATTCCTATTCTTTTCATCAATAAATAATGCAATCTTCT 120  
DB 61 GCTATTAGCACTAAATGCAATTCCTATTCTTTTCATCAATAAATAATGCAATCTTCT 120

QY 121 ACAGGAGTAACCTTAAACAGGCTTAACGGTCAATATCAATTTGAAACACTATTCCTGTA 180  
DB 121 ACAGGAGTAACCTTAAACAGGCTTAACGGTCAATATCAATTTGAAACACTATTCCTGTA 180

QY 181 AGTAATGAAATGCTGCTGTTGCTAAATCTCTCCAGGAGCTATTTTATTAAGAGTACT 240  
DB 181 AGTAATGAAATGCTGCTGTTGCTAAATCTCTCCAGGAGCTATTTTATTAAGAGTACT 240

QY 241 TTTTATTAATTAATTTTCAAGTTTGGCAGATATTAGTATTAATTTAAAGAAATGAA 300  
DB 241 TTTTATTAATTAATTTTCAAGTTTGGCAGATATTAGTATTAATTTAAAGAAATGAA 300

QY 301 CAACACCAAGTCTTTTAAACAGGCTTAATCAGAGTATACCTTTAAAGGAAAGATGT 360  
DB 301 CAACACCAAGTCTTTTAAACAGGCTTAATCAGAGTATACCTTTAAAGGAAAGATGT 360

QY 361 GACGAGTATCCTGCTTACAGAAAGTATCAACAGAAATCCTTTTGATTTTAAAAA CAAA 420  
DB 361 GACGAGTATCCTGCTTACAGAAAGTATCAACAGAAATCCTTTTGATTTTAAAAA CAAA 420





QY 541 GACTCTCATCGTATGAGCCAGGTTTAACTTAATCACTTTGGACAACTTTCAGCAGATTGTGATG 600  
Db 2282 GACTCTCATCGTATGAGCCAGGTTTAACTTAATCACTTTGGACAACTTTCAGCAGATTGTGATG 2341  
QY 601 GTAGTTCTTCCAAGTAAATCTTTGAGAGAAATTTTCAGCAGTATTTTACAGATGATATTGAG 660  
Db 2342 GTAGTTCTTCCAAGTAAATCTTTGAGAGAAATTTTCAGCAGTATTTTACAGATGATATTGAG 2401  
QY 661 ACCGTTGAGGTATTTTCTCACCAGCCAAATCTTTGTTTTCAGAGTGAACACATTTCTTTT 720  
Db 2402 ACCGTTGAGGTATTTTCTCACCAGCCAAATCTTTGTTTTCAGAGTGAACACATTTCTTTT 2461  
QY 721 TATACACGCTCTTGAAGGAAATATCCCGATACAGACCGTTTATTAATGACAGAAATTT 780  
Db 2462 TATACACGCTCTTGAAGGAAATATCCCGATACAGACCGTTTATTAATGACAGAAATTT 2521  
QY 781 GAGCGGAGGTGTTTTCACCAACCAATCCCTTCGCCACGCTATGGAACGTCCTTCTTG 840  
Db 2522 GAGCGGAGGTGTTTTCACCAACCAATCCCTTCGCCACGCTATGGAACGTCCTTCTTG 2581  
QY 841 ATTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTGAGATTACTCAAAATCATATTTCA 900  
Db 2582 ATTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTGAGATTACTCAAAATCATATTTCA 2641  
QY 901 GCTCATGTTAACTCACTGAGGTTGCTAAGGTAACGAGGATTTAGATATTGTTAGTCAG 960  
Db 2642 GCTCATGTTAACTCACTGAGGTTGCTAAGGTAACGAGGATTTAGATATTGTTAGTCAG 2701  
QY 961 TCTGTAGTGATTAACTATCAGCTTCAATCAACTTACCTTATTTAGCTCTTTAAAGCT 1020  
Db 2702 TCTGTAGTGATTAACTATCAGCTTCAATCAACTTACCTTATTTAGCTCTTTAAAGCT 2761  
QY 1021 ATTAAAGTGAACAGTAAATTCATTTCTTATCACCAGTTCGACCATTCACCCTAACA 1080  
Db 2762 ATTAAAGTGAACAGTAAATTCATTTCTTATCACCAGTTCGACCATTCACCCTAACA 2821  
QY 1081 CCAGCGGATGAGGAGAAAGTTTATCCAAATTAATTAACCAAGTACGACCAAC 1134  
Db 2822 CCAGCGGATGAGGAGAAAGTTTATCCAAATTAATTAACCAAGTACGACCAAC 2875

RESULT 4

CP000003-00  
WPCOMMENT

Sequence split into 19 fragments LOCUS CP000003 Accession CP000003

Fragment Name	Begin	End
CP000003_00	1	110000
CP000003_01	100001	210000
CP000003_02	200001	310000
CP000003_03	300001	410000
CP000003_04	400001	510000
CP000003_05	500001	610000
CP000003_06	600001	710000
CP000003_07	700001	810000
CP000003_08	800001	910000
CP000003_09	900001	1010000
CP000003_10	1000001	1110000
CP000003_11	1100001	1210000
CP000003_12	1200001	1310000
CP000003_13	1300001	1410000
CP000003_14	1400001	1510000
CP000003_15	1500001	1610000
CP000003_16	1600001	1710000
CP000003_17	1700001	1810000
CP000003_18	1800001	1899877

LOCUS CP000003 1899877 bp DNA circular BCT 03-AUG-2004

DEFINITION Streptococcus pyogenes MGAS10394, complete genome.

ACCESSION CP000003

VERSION CP000003.1 GI:50902420

KEYWORDS

SOURCE Streptococcus pyogenes MGAS10394

ORGANISM Streptococcus pyogenes MGAS10394

Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

Streptococcus.  
1 (bases 1 to 1899877)  
Banks,D.J., Porcella,S.F., Barbican,K.D., Beres,S.B., Philips,L.E.,  
Voyich,J.M., DeLeo,F.R., Martin,J.M., Somerville,G.A. and  
Musser,J.M.  
Progress toward Characterization of the Group A Streptococcus  
Metagenome: Complete Genome Sequence of a Macrolide-Resistant  
Serotype M6 Strain  
J. Infect. Dis. 190 (4), 727-738 (2004)  
15272401  
2 (bases 1 to 1899877)  
Banks,D.J., Porcella,S.F., Barbican,K.D., Beres,S.B., Philips,L.E.,  
Voyich,J.M., DeLeo,F.R., Martin,J.M., Somerville,G.A. and  
Musser,J.M.  
Direct Submission  
Submitted (20-JUL-2004) Laboratory of Human Bacterial Pathogenesis,  
Rocky Mountain Laboratories, National Institute of Allergy and  
Infectious Disease, National Institutes of Health, 903 S. 4th  
Street, Hamilton, MT 59840, USA  
Location/Qualifiers  
1..1899877  
/organism="Streptococcus pyogenes MGAS10394"  
/mol\_type="genomic DNA"  
/strain="MGAS10394"  
/serotype="M6"  
/db\_xref="taxon:286636"  
202..1557  
/locus\_tag="M6\_Spy0001"  
202..1557  
/locus\_tag="M6\_Spy0001"  
/note="Dnaa"  
/codon\_start=1  
/transl\_table=11  
/product="chromosomal replication initiator protein"  
/protein\_id="AAT86136.1"  
/db\_xref="GI:50902421"  
/translation="MTENEQIFWNRVLELAQSOLKQATYEFFVHDARLLKVDKHIATI  
YLDQKELFWKNLKDVLITAGFEVYNAQISVDYFEEELMEQNTKINQPKQAL  
NSLPTVSDLSKYSFENIQGDNENWAASIAVANTPGTYYNPLFWGGFGLKTH  
LLAIGNSVLLENPNARIKYITAEFNIFEVHIHRLDTMDELKERFLNLDLLIIDIQ  
SLAKTSLSGTOEEFFNTFNALNNKQIVLTSRTPDHLNDELRLVTRFKWGLTVNI  
TPDPFETRAVAILTKIOEYFIIPQDTIEYLAGOFDSNVRLGALKDISLVANFKOI  
DTTVIDAAEAIARAKQDGPQKTVPIEEIOAGYKGYVTVEKIKATKRTQNVILAR  
QVAMFLAREMTDNLSPKIGKEFGGRDHSVLHAYNKIKNMISQDESLEIETIKNKI  
K"  
1712..2848  
/locus\_tag="M6\_Spy0002"  
1712..2848  
/locus\_tag="M6\_Spy0002"  
/EC\_number="2.7.7.7"  
/codon\_start=1  
/transl\_table=11  
/product="DNA polymerase III beta chain"  
/protein\_id="AAT86137.1"  
/db\_xref="GI:50902422"  
/translation="MIQFSINRITLFIHALNATKRAISTKNAIPILSSIKIEVTPRGVT  
LTGNSQISIENTIPVSNENAGLLITPGAILLEAFINIISSLPDISINVEIEQH  
QVLTSGKSEITLKQVDQYPRQEVSTENPLIKTKLKSIIAETAFASLQESRP  
ILTGHVLSNHKDFKAHSIDYSHRMSORLITLNTSADPDVVPSPKREFSAFTDD  
IETVEFSPSQIILFRSEHISFYTRLLEGNYPDTRLMTFETEFVNTQSLRHAME  
RAFLISATQNGTVKLEITQNHISAHVNSPEVGKVNEDLDIVSQSGSLTISFNPYIL  
IESUKAIKSETVKIHIFUSPVRPFTLTTPGDEESFIQLITPVRTN"  
2923..3120  
/locus\_tag="M6\_Spy0003"  
2923..3120  
/locus\_tag="M6\_Spy0003"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical cytosolic protein"  
/protein\_id="AAT86138.1"  
/db\_xref="GI:50902423"  
/translation="MYQIGSFVEMKKPHACVIKETGKKANQMKVLRVGADIKIQCTNC  
QHVIMSRDYDFDRKKVLQP"

```

gene
3450..4565
/locus_tag="M6_Spy0004"
CDS
3450..4565
/locus_tag="M6_Spy0004"
/note="probable translation factor"
/codon_start=1
/transl_table=11
/product="GTP-binding protein"
/protein_id="AAT86139.1"
/db_xref="GI:50902424"
/translation="MALTAGIVGLPNVKGSTLFLNATKAGAAANYPFATIDPNVGMV
EVPDRLOKLTETLTPKTVPTTEFTDIAGIVKGSRGEGLNKFLANIREVDIAIV
VSPADENVMREQGEDAFVPIADIDITINLELILADLESINKRYARVERKWARQVH
KSAADFNVQKIPVLEDGKGARTIEFTEBEAKVVKGLFTTTPFLVLYANVDKDV
ANPDGDIYVQKIRDFAAENAEVVISARABEESILEDDEKEEFLEAIGLTPESGVDK
LTRAAYHLGLGTFTAGEKEVRAMTFKRGIKAPOAAGIIHSDFERGFIRAVTMSYDD
LMIYSEKAVKAGRLREGEKGVVQDGDIMEFRNV"
4590..5204
/locus_tag="M6_Spy0005"
CDS
4590..5204
/locus_tag="M6_Spy0005"
/EC_number="3.1.1.29"
/codon_start=1
/transl_table=11
/product="Peptidyl-tRNA hydrolase"
/protein_id="AAT86140.1"
/db_xref="GI:50902425"
/translation="MEESHSPFGILERNVMKIVGLGNPGSKYBKTKNIGFMAIDN
IVNLDVFTDDNFKAQIGSTFINHEKVYFVKPTTFMNNSGIAVKALLTYNIDITD
LVIYDDIMEYKSLRLSKSGAGHGKISIIAHICTOENRINKVIGRPLKGMTVI
SHVMQFNTEDNIALSLTLDVVNAVVKYLQNDPEKTMQKFG"
5207..8710
/locus_tag="M6_Spy0006"
CDS
5207..8710
/locus_tag="M6_Spy0006"
/codon_start=1
/transl_table=11
/product="transcription-repair coupling factor"
/protein_id="AAT86141.1"
/db_xref="GI:50902426"
/translation="MDILELFSQNKVQSHSGTLTLGRQLVMGLSGSSKALATASAV
LDDQKIVVSTQNEVKLSADLSLDEELVFOFPADDDAAAEFTFASMDKALSRI
ETLQPLRNPQSVGLIVLSLRLTLLPNPDTVKSOIOLTWGEDYDSDLTKQLMTIG
YKQSVQVSPGEFSPRGDILDIYEITQPLPYRLEFPFGDDISIQPFPETOKSEFQLE
GIFINPADDLIFEASDPORGIEQLEKALQADDDKSYLEDVLAVSKNGFKHKDIRF
QSLFYKEWSLSDIYIPKGTPIFFDDFKLVKNARFDLEIANLLTDLQOQKLSNLN
IFADNYRELHYKATPFNSPHKGLNFKPDQMHQLTQYAMQEFNFQPLIIDIKRY
QKNQTVIVQVESQVAYERLEKSFQDYQFRLPLVSANOIVRSQIVIGATSSGFYPA
DEKALIIHEIYHKIKRRARRNINSABRLKDYNEIAGDYVYVHVHVGIGRFLGLE
TIOGIRHDVVTIOYONSDRISLPIDQISLSKYVSADGKEPKINKLNDRFOKTKQ
KVAREQVEDLKLKLYARSQOQGFSPDDDLQAPDDDFAFVETEDQLRSIKEIK
ADMSMQMDRLLDVYDVGFKTEVAMRAAFKAVNDHKQVAVLVTTLAQOQHYENFKA
RFENYPVEVDVLSRFSKKEQAETLERVQKQDIIIGTHRLSKDVVFDLGLIVID
BEQFVGNKHELTKEKTKVDVLTATPIPTLHSMGLGIRLSLVETPPTNRYPVQ
TYVLNPGVLREAIIREMDRGGQIFVYVNVKVDITIEKKVARELQELVPEASIGFVHGQ
SEIOENTLIDFINGVDVLTATIIETGVDISUNVTLEFINADHMGSLTLYQLRGVR
GRSRIAYAVLMVDPKLTIEVSEKRLKALKGTELSGFRKIAMRDISIRYAGNILGA
SQSGFIDSVGEMYSQILLEQIAKQKQTVRQKGTNEINQIDAYLPDDYIADERQK
IDITYKIREISRESDYNLQDELIDREFEYDPQVAYLLEALLKHYMDNAPAEIIVERK
NQVIRFEVTSLTFLTYQDYFALS KTHLKAKI SEHQGKIDIVFDVRHQDYRILBE
LMLFGERLSEIKIRKNNSVFK"
8848..9144
/locus_tag="M6_Spy0007"
CDS
8848..9144
/locus_tag="M6_Spy0007"
/codon_start=1
/transl_table=11
/product="Heat shock protein 15"
/protein_id="AAT86142.1"
/db_xref="GI:50902427"
/translation="MKRIIVEIMRLDKYLKSVRLIKRRSVAKEVADKGRIVKNGILAK
SSINVKLNHIEISFGNKLITVRVIEIKDKSKEDALKMYEIISETRITLNEBA"
9131..9502

```

```

/locus_tag="M6_Spy0008"
9131..9502
/locus_tag="M6_Spy0008"
/note="cell division protein"
/codon_start=1
/transl_table=11
/product="DivIC"
/protein_id="AAT86143.1"
/db_xref="GI:50902428"
/translation="MKKPSIVOLNHHYIKKENLKKKFFEEBSOKNRNFMGWLVSMMF
LFIPTYNLKSVDYDFERQNOQVVKLKKKEYNELSKTKKERQLAERLKKDDNFVKYAR
AKYLSREGEMIIYPIGLLPK"
9499..9624
/locus_tag="M6_Spy0009"
CDS
9499..9624
/locus_tag="M6_Spy0009"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAT86144.1"
/db_xref="GI:50902429"
/translation="MMMDNIIKKIEAFIASFDDKKLAEQENQKVKEESQYIQKN"
9637..10923
/locus_tag="M6_Spy0010"

```

```

Query Match
Best Local Similarity 98.6%; Score 1118; DB 1; Length 110000;
Matches 1124; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
Qy 1 ATGATTCAATTTTCAATTAATCGCACATTAATTCATGCTTTAAATGCACTAAACGT 60
Db 1712 ATGATTCAATTTTCAATTAATCGCACATTAATTCATGCTTTAAATGCACTAAACGT 1771
Qy 61 GCTATTAGCACTAAATGCCATTCCTATTCTTTCATCAATAAATAATGAAGTCACCTCT 120
Db 1772 GCTATTAGCACTAAATGCCATTCCTATTCTTTCATCAATAAATAATGAAGTCACCTCT 1831
Qy 121 ACAGGAGTAACTTTAAACAGGGTCTAACGGTCAATATCAATTTGAAACACATTTTCCCTGTA 180
Db 1832 ACAGGAGTAACTTTAAACAGGGTCTAACGGTCAATATCAATTTGAAACACATTTTCCCTGTA 1891
Qy 181 AGTAATGAAATGCTGTTTGTCTTAATTAATCTCTCCAGAGCTATTTTATTAGAAGCTAGT 240
Db 1892 AGTAATGAAATGCTGTTTGTCTTAATTAATCTCTCCAGAGCTATTTTATTAGAAGCTAGT 1951
Qy 241 TTTTATTAAATATTTTCAAGTTTCCAGATATTTAGTATAATGTTAAAGAAATTTGAA 300
Db 1952 TTTTATTAAATATTTTCAAGTTTCCAGATATTTAGTATAATGTTAAAGAAATTTGAA 2011
Qy 301 CAACACCAAGTTGTTTAAACAGTGGTAAATCAGAGATTACCTTTAAAGGAAAAGATGTT 360
Db 2012 CAACACCAAGTTGTTTAAACAGTGGTAAATCAGAGATTACCTTTAAAGGAAAAGATGTT 2071
Qy 361 GACCAAGTATCTCGTCTACAGAGATATCAACAGAAATCTCTTTGATTTTAAAAACAAA 420
Db 2072 GACCAAGTATCTCGTCTACAGAGATATCAACAGAAATCTCTTTGATTTTAAAAACAAA 2131
Qy 421 TTATTGAAGTCTATTATTGCTGAAAACAGCTTTTGCAGCCAGTTTCAAGAAAGCTCGCTCT 480
Db 2132 TTATTGAAGTCTATTATTGCTGAAAACAGCTTTTGCAGCCAGTTTCAAGAAAGCTCGCTCT 2191
Qy 481 ATTTTAAACAGGAGTTTCAATTTGATTAAGTAATCATATAAGATTTTAAAGCAGTAGCGACT 540
Db 2192 ATTTTAAACAGGAGTTTCAATTTGATTAAGTAATCATATAAGATTTTAAAGCAGTAGCGACT 2251
Qy 541 GACTCTCATCGTATGACCAAGCTTTTAATCACTTTGACAAATCTTTCAGCAGATTTTATG 600
Db 2252 GACTCTCATCGTATGACCAAGCTTTTAATCACTTTGACAAATCTTTCAGCAGATTTTATG 2311
Qy 601 GTAGTTCTTCCAAAGTAAATCTTTGAGAGAAATTTTACAGAGTATTTACAGATGATTTAG 660
Db 2312 GTAGTTCTTCCAAAGTAAATCTTTGAGAGAAATTTTACAGAGTATTTACAGATGATTTAG 2371

```









QY 781 GAGCGGAGGTGTTTCAATACCAATCCCTTCGCCAGCTATCGAAGTGCCTCTTGTG 840  
Db 2489 GAGCGGAGGTGTTTCAATACCAATCCCTTCGCCAGCTATCGAAGTGCCTCTTGTG 2548  
QY 841 ATTTCCTAATGCTACTCAAAATCGTACTGTTAAGCTTTGAGATTACTCAAAATCATATTCA 900  
Db 2549 ATTTCCTAATGCTACTCAAAATCGTACTGTTAAGCTTTGAGATTACTCAAAATCATATTCA 2608  
QY 901 GCTCATGTTAATCACTACCTGAGTTCGGTAAAGTAAACGAGGATTAGATTGTTAGTCAG 960  
Db 2609 GCTCATGTTAATCACTACCTGAGTTCGGTAAAGTAAACGAGGATTAGATTGTTAGTCAG 2668  
QY 961 TCTGCTAGTGAATTAACCTATCACTACCTCAATCAACTTACCTTATGAGTCTTTAAAGCT 1020  
Db 2669 TCTGCTAGTGAATTAACCTATCACTACCTCAATCAACTTACCTTATGAGTCTTTAAAGCT 2728  
QY 1021 ATTTAAAGTGAACAGTAAATTTTCAATTTCTTATCACCAGTTCGACCATTCACCCCTAACA 1080  
Db 2729 ATTTAAAGTGAACAGTAAATTTTCAATTTCTTATCACCAGTTCGACCATTCACCCCTAACA 2788  
QY 1081 CCAGCGGATGAGGAAGAGTTTTCATTAATTAATTAACACCAAGTACGACCAAAAC 1134  
Db 2789 CCAGCGGATGAGGAAGAGTTTTCATTAATTAATTAACACCAAGTACGACCAAAAC 2842  
RESULT 6  
AE014136 51896 bp DNA linear BCT 19-JUL-2002  
LOCUS Streptococcus pyogenes MGAS315, section 1 of 37 of the complete  
DEFINITION genome.  
ACCESSION AE014136 AE014074  
VERSION AE014136.1 GI:21903716  
KEYWORDS  
SOURCE  
ORGANISM Streptococcus pyogenes MGAS315  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE 1 (bases 1 to 51896)  
AUTHORS Beres,S.B., Sylva,G.L., Barbican,K.D., Lei,B., Hoff,J.S.,  
Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,  
Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and  
Muser,J.M.  
TITLE Genome sequence of a serotype M3 strain of group A Streptococcus:  
Phage-encoded toxins, the high-virulence phenotype, and clone  
emergence  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)  
REFERENCE 2 (bases 1 to 51896)  
AUTHORS Beres,S.B., Sylva,G.L., Barbican,K.D., Lei,B., Hoff,J.S.,  
Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,  
Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and  
Muser,J.M.  
TITLE Direct Submission  
JOURNAL Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis,  
Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St.,  
Hamilton, MT 59840, USA  
FEATURES  
source  
1. 51896  
/organism="Streptococcus pyogenes MGAS315"  
/mol\_type="genomic DNA"  
/strain="MGAS315"  
/serotype="M3"  
/db\_xref="taxon:198466"  
/note="group: A"  
232. .1587  
/gene="dnaA"  
/note="synonym: Spym3\_0001"  
232. .1587  
/gene="dnaA"  
/note="best blastp hit: gb|AAF71535.1|AF255728.1  
(AF255728) chromosomal initiator protein DnaA  
[Streptococcus pyogenes]"  
/codon\_start=1  
/transl\_table=11

/product="chromosomal initiator protein"  
/protein\_id="AAM78608.1"  
/db\_xref="GI:21903717"  
/translation="MTENEQIFMNRVLELAQSLKQATYEFFVHDARLLKVDKHIATI  
YLDQKELFWENLKVLTITAGFVYNAQISVDVVFEDLMIEQNOTKINOKPOQAL  
NSLPTVSDLSKYSFENFIQODENRWAASIAVANTPGTTPFLTWGPGGLKTH  
LNAIGSVLLENFARIKYTAENFNEFVHRLDTMDELKEFRNLDDLLLDIDQ  
SLAKTUSGTOEEFFFNALHNNNKQILVTSDRTPDHLNDELDELVTRFKWGLTVNI  
TPPDFTRVAILTNKIQYENFIFFQDTIEYLAGQFDSNVRDLLEGALKDISLVANFKQI  
DTITVDAEAIRAKODGPRWTPIBEEIQAGQVGFYGVTKETKATKQTQNIVLAR  
QVAMFLAREMTDNLSPKIGKEFGCRDHSVTLHAVNKIKNMTSQDSLEIETIKNKI  
K"  
1742. 2878  
/gene="dnaN"  
/note="synonym: Spym3\_0002"  
1742. 2878  
/gene="dnaN"  
/note="best blastp hit: gb|AAK33147.1| (AE006472) beta  
subunit of DNA polymerase III [Streptococcus pyogenes M1  
GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="beta subunit of DNA polymerase III"  
/protein\_id="AAM78609.1"  
/db\_xref="GI:21903718"  
/translation="MIQFSINRTLFIHALNATKRAISTYNAIPILSSIKIEVTPGT  
LTGSGQISIENTIPVSNENAGLLITSPGAILLEASFIIISLPDININVEIQH  
QVLTISGKSEITLKGQDVQYPRQVSTENPLILKTLKLSIIAETAFASLQBSRP  
ILTGVIHLSNHDKFAVATDSHRMSORLITDNTADFVVIPOKSLREFSAVFTDD  
IETVEFSPSQILFRSEHISFYTRLLEGNYPDTRLMTFETEVVENTOSLRHAME  
RAFLSNATQNGTVKLEITQNHISAHVNSPEVGKVNEDLDIVSGSGLTISFNPTYL  
IESLKAIKSETVKIHLFSPVRPFTLTGDBEESFIQLITPVRTN"  
2953. 3150  
/gene="Spym3\_0003"  
2953. 3150  
/gene="Spym3\_0003"  
/note="best non-GAS blastp hit: gb|AAC45338.1| (AF000658)  
ORF [Streptococcus pneumoniae]"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAM78610.1"  
/db\_xref="GI:21903719"  
/translation="MYQIGSPVEMKKPHACVIKETGKANQWKLVRGADIKIQCTNC  
QHVIMSGRYDFERKLKKVLP"  
3480. 4595  
/gene="Spym3\_0004"  
3480. 4595  
/gene="Spym3\_0004"  
/note="best blastp hit: gb|AAK33149.1| (AE006472) putative  
GTP-binding protein [Streptococcus pyogenes M1 GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="putative GTP-binding protein"  
/protein\_id="AAM78611.1"  
/db\_xref="GI:21903720"  
/translation="MALTAGIVGLPNVGKSTLFNAITKAGAAANYPFATIDPNVGMV  
EVPDERLQKLTELITPKKTVPTTFDITAGIVKAGSGEGLNKFLANIREVDAIVH  
VVRAPDDENVMREQRDAFVDVADIDTINLELILADLESINKRYARVEKMAKTQKD  
KESVAEFNVLOKIKPVEDGKSARTIETDEAKVKGLFLLTKPVLVYANVDEKDV  
ANPDGIDYKQIRDFATENAENVVISARAEELLEDDEDEKEFLEAIGLTESGVDK  
LTRAAYHLGLGTYFTAGEKEVRWTFRGIKAPQAAGLIHSDFERGFIRAVTMSYDD  
LMTYSSEKAVKEAGRLREEGKEYVYVQGDIMEFRNV"  
4665. 5234  
/gene="pth"  
/note="synonym: Spym3\_0005"  
4665. 5234  
/gene="pth"  
/note="best blastp hit: gb|AAK33150.1| (AE006472) putative  
peptidyl-L-trna hydrolase [Streptococcus pyogenes M1 GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="putative peptidyl-L-trna hydrolase"

```
/protein_id="AAW78612.1"
/db_xref="GI:21903721"
/translation="MYKMIYVIGLGNPGSKVEKTKYHNGMAIDNIVKNDVFTTDDKNF
KLGISGTFINHEKVCEVKEFTFMNNSGIAVXALTYINIDITDLIVYDDIDMEVSKL
RQSGKSAGHNGIKSIIAHIGTQEFNRIKQIGRPLKGMTVINHVWGQFNTEDNIAI
SUTLDRVNAVFXILOENDFEKTMQKFG"
gene
5321..8740
/genes="trcf"
/notes="synonym: Spym3_0006"
CDS
5321..8740
/genes="trcf"
/notes="best blastp hit: gb|AAK33151.1| (AE006472) putative
transcription-repair coupling factor [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative transcription-repair coupling factor"
/protein_id="AAW78613.1"
/db_xref="GI:21903722"
/translation="WGLSGSGSKALAIASAYLDQKKIVVTVSTONEVEKLADLSLL
DELVFFQFADVAABEFIFASMDKALRIETQLFNRPKSQGVLIYLSGLRTLLPN
PDVFTKSIQIQTGVGDDSDTLTKQMTIGTVQKVSQVISPGESESGDILDIEYIQE
LPYLEFEGDDIDSIROFHPETOKSPOLSGVINPASDLIFASDFQKIQSLEKAL
QTAQDKKSYLVDVLAIVKNGFKHDIRKQSLFYKESWLLDIIPKGPPIFFDFQK
LVDKARFDEIANLPTDILQCKALSNLNFADNIRELHYKPAFFPFHFKGLGNI
KFDOMHOLTOYQAFQEPNFPFLIDEIKRYKQNTTVIVQVESQYAYERLEKSFQDYQ
FRPLVSAHQIVSRESQIVIGASSGFYFADEKALITEHYHKKIKRARRSNISN
AERLQDNEAVGDDVYVHNGHGFLEGTIETIQIQRHYVTVIOYNSDRISLPIQ
ISSLKVSVDGKEPKINLNGRFQKQKVARQVEDIADLLKLAERSQKQSF
SPDDLPQAFDDPAFVETEDLSRKEIKADMESQPMQDRLLGVGDFGKTEVAMRA
AFKAVNDHKQAVLPTVLAQOHYENFKARPENYFVVDLSRFSKKEQATLERV
RKQDIIITGHELLSKDVPSDLGLVIDESQFVGKHKTLEKTKVDVLTAT
PIPTLHMSGLRDLISVETPFTNRPVQTVLENNPGLVREAIIREMDRGQIFVY
YKVDIETKVAELSGVPEASIFVHGQMSIOLENTLIDFINGDYDLVLTAT
GVDSINVTLIFENADHMLSTLYQLRGVRSNRIFAYAYMYRDPKVLTEVSKRLE
AIGFELSGSFKAMRDLISIRGAGNLCASQSGRIASVGMYSQQLLEQAIASQKQ
TTPVQKNTENINQIDAYLPDDVIADEROKIDIKYRIEIOSREDYLNQDELDQFG
EYDQVAYLLEIALKLYMNDPAFELVERKKNQIVRFEVLSLSFLQDYFEALSKT
HLKAKISEHQKIDIVFEDVRHQDRIILEELMLFGERLSEIKIRKQNSVFK"
gene
8902..9174
/genes="Spym3_0007"
/notes="synonym: Spym3_0007"
/notes="best non-GAS blastp hit: gb|AAK74200.1| (AE007318)
S4 domain protein [Streptococcus pneumoniae TIGR4]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAW78614.1"
/db_xref="GI:21903723"
/translation="WRLDKYLKVSRLIKRRSVAKEVADKGRIVKNGILAKSSTNINL
DHEISFGNKLTVRVIEIKDSTKEDALKWYELISETRITLNEA"
gene
9161..9532
/genes="divic"
/notes="synonym: Spym3_0008"
/notes="best blastp hit: gb|AAK33153.1| (AE006472) putative
cell division protein (divic) [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="putative cell division protein"
/protein_id="AAW78615.1"
/db_xref="GI:21903724"
/translation="MKKPSIVQLNNHYIKKENLKKXFEESQKRNFMGWLIVSNMF
LFILPTYNLVKSYVDFEKQOQVVKLKEYNELSKTKKEKQLAERLKDDNFVKIYAR
9667..10953
/genes="Spym3_0009"
/notes="synonym: Spym3_0009"
/notes="best non-GAS blastp hit: gb|AAK98813.1| (AE008385)
S4 domain protein [Streptococcus pneumoniae TIGR4]"
/codon_start=1
/transl_table=11
/product="putative cell division protein"
/protein_id="AAW78616.1"
/db_xref="GI:21903725"
/translation="MRKLLAAMLMFTFFLTPLVISTEKKLIFSKNVAVQLKQVVQST
```

```
Conserved hypothetical protein [Streptococcus pneumoniae
R6]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAW78616.1"
/db_xref="GI:21903725"
/translation="MRKLLAAMLMFTFFLTPLVISTEKKLIFSKNVAVQLKQVVQST
```

```
Query Match 98.3%; Score 1114.8; DB 1; Length 51896;
Best Local Similarity 98.9%; Pred. No. 2.2e-195;
Matches 1122; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 1 ATGATTCAATTTTCAATTAATCGCACATTTATTATTCATGCTTTAAATAACAATTAACGT 60
DB 1742 ATGATTCATTTTCAATTAATCGCACATTTATTATTCATGCTTTAAATAACAATTAACGT 1801
QY 61 GCTATTAGCATTAAATGCGCATTTCTTATTCATCAATTAATAATTTGAAGTCACCTCT 120
DB 1802 GCTATTAGCATTAAATGCGCATTTCTTATTCATCAATTAATAATTTGAAGTCACCTCT 1861
QY 121 ACAGGAGTAACCTTTAAACAGGCTCTAACGGTCAATATCAATTTGAAACACATTTCTCTGA 180
DB 1862 ACAGGAGTAACCTTTAAACAGGCTCTAACGGTCAATATCAATTTGAAACACATTTCTCTGA 1921
QY 181 AGTAATGAAATGCTGTTTGTCTTAAATACCTCTCCAGGAGCTATTTTATTAGAGCTAGT 240
DB 1922 AGTAATGAAATGCTGTTTGTCTTAAATACCTCTCCAGGAGCTATTTTATTAGAGCTAGT 1981
QY 241 TTTTATTATTAATTTTCAAGTTTGCAGATATTAGTATTAATTTTAAAGAAATTTGAA 300
DB 1982 TTTTATTATTAATTTTCAAGTTTGCAGATATTAGTATTAATTTTAAAGAAATTTGAA 2041
QY 301 CAACACCAAGTTGTTTAAACAGTGTGTAATCAGAGATTACCTTTAAAGAAATTTGAA 360
DB 2042 CAACACCAAGTTGTTTAAACAGTGTGTAATCAGAGATTACCTTTAAAGAAATTTGAA 2101
QY 361 GACAGTATCCCTGCTCTACAGAGATATCAACAGAAATCCTTTGATTTTAAACACAAA 420
DB 2102 GACAGTATCCCTGCTCTACAGAGATATCAACAGAAATCCTTTGATTTTAAACACAAA 2161
QY 421 TTATTGAAGTCTATTATTCGCTGAAAACAGCTTTTGCAGCCAGTTTACAGAGATTTGATG 480
DB 2162 TTATTGAAGTCTATTATTCGCTGAAAACAGCTTTTGCAGCCAGTTTACAGAGATTTGATG 2221
QY 481 ATTTTAAACAGGAGTTCATATTTGATTAAGTATCAATAAGATTTTAAAGCAGTAGCGACT 540
DB 2222 ATTTTAAACAGGAGTTCATATTTGATTAAGTATCAATAAGATTTTAAAGCAGTAGCGACT 2281
QY 541 GACTCTCATCGTATGAGCAACAGTTTAAATCACTTTGGACAATCTTCAGCAGATTTGATG 600
DB 2282 GACTCTCATCGTATGAGCAACAGTTTAAATCACTTTGGACAATCTTCAGCAGATTTGATG 2341
QY 601 GTAGTCTCTCCAGTAAATCTTTGAGAGATTTTTCAGAGATTTTACAGATGATATTGAG 660
DB 2342 GTAGTCTCTCCAGTAAATCTTTGAGAGATTTTTCAGAGATTTTACAGATGATATTGAG 2401
QY 661 ACCGTGTAGGATTTTCTCACCAGCAATCTTTGAGAGATTTTTCAGAGATTTTACAGATGATATTGAG 720
DB 2402 ACCGTGTAGGATTTTCTCACCAGCAATCTTTGAGAGATTTTTCAGAGATTTTACAGATGATATTGAG 2461
QY 721 TATACACGCTCTTAAAGAGAAATTTCCCGATACAGACCGTTTATTAATGACAGAAATTT 780
DB 2462 TATACACGCTCTTAAAGAGAAATTTCCCGATACAGACCGTTTATTAATGACAGAAATTT 2521
QY 781 GAGACGAGGTTTGTTTTCAATACCCATCCCTTCGCCACGCTATGAACTGCTCTCTTG 840
DB 2522 GAGACGAGGTTTGTTTTCAATACCCATCCCTTCGCCACGCTATGAACTGCTCTCTTG 2581
QY 841 ATTTCTAATGCTACTCAAAATGCTAGTTTAAAGCTTTGAGATTTACTCAAAATCATATTTCA 900
DB 2582 ATTTCTAATGCTACTCAAAATGCTAGTTTAAAGCTTTGAGATTTACTCAAAATCATATTTCA 2641
```

QY 901 GCTCATGTTAACTCACCCTGAGGTGCTAAGGTAACGAGGATTAGATATTGTTAGTCAG 960  
Db 2642 GCTCATGTTAACTCACCCTGAGGTGCTAAGGTAACGAGGATTAGATATTGTTAGTCAG 2701  
QY 961 TCTGCTAGTGAATTAACATATCAGCTTCAATCAACTTACCTTATTGAGTCTTTAAAGCT 1020  
Db 2702 TCTGCTAGTGAATTAACATATCAGCTTCAATCAACTTACCTTATTGAGTCTTTAAAGCT 2761  
QY 1021 ATTAAGAAGTGAACAGTAAATTCATTTCTTATCAGCTTCACTTACCTTACCTTAAAC 1080  
Db 2762 ATTAAGAAGTGAACAGTAAATTCATTTCTTATCAGCTTCACTTACCTTACCTTAAAC 2821  
QY 1081 CCAGCGCATGAGGAAGAGTATTCATCAATTAATTAACACAGTACGACCAAC 1134  
Db 2822 CCAGCGCATGAGGAAGAGTATTCATCAATTAATTAACACAGTACGACCAAC 2875

RESULT 7  
AP005141  
LOCUS AP005141 311600 bp DNA linear BCT 07-APR-2004  
DEFINITION Streptococcus pyogenes SSI-1 DNA, complete genome, section 1/6.  
ACCESSION AP005141 BA000034  
VERSION AP005141.1 GI:28810157  
KEYWORDS Streptococcus pyogenes SSI-1  
SOURCE Streptococcus pyogenes SSI-1  
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.

1  
Nakagawa, I., Kurokawa, K., Yamashita, A., Nakata, M., Tomiyasu, Y.,  
Okahashi, N., Kawabata, S., Yamazaki, K., Shiba, T., Yasunaga, T.,  
Hayashi, H., Hattori, M. and Hamada, S.  
Genome sequence of an M3 strain of Streptococcus pyogenes reveals a  
large-scale genomic rearrangement in invasive strains and new  
insights into phage evolution  
Genome Res. 13 (6A), 1042-1055 (2003)

JOURNAL 2 (bases 1 to 311600)  
MEDLINE 22683278  
PubMed 12799345  
REFERENCE Hattori, M., Yamashita, A., Oshima, K. and Shiba, T.  
Direct Submission  
AUTHORS Submitted (01-MAY-2002) Masahira Hattori, Kitasato Institute for  
TITLE Life Sciences, Kitasato University, Kitasato 1-15-1, Sagamiara,  
JOURNAL Kanagawa 228-8555, Japan  
(E-mail: hattori@genome.is.kitasato-u.ac.jp,  
URL: http://genome.is.kitasato-u.ac.jp/, Tel: 81-42-778-8194,  
Fax: 81-42-778-8193)

COMMENT This work was done in collaboration with Ichiro Nakagawa, Ken  
Kurokawa, Masanobu Nakata, Yusuke Tomiyasu, Nobuo Okahashi,  
Shigetada Kawabata, Kiyoshi Yamazaki, Teruo Yasunaga, Shigeyuki  
Hamada (Osaka University), Hideo Hayashi (University of Tsukuba),  
and supported by the Research for the Future Program of the Japan  
Society for the Promotion of Science. This clone was isolated from  
a patient presenting with toxic shock like syndrome.

FEATURES  
Location/Qualifiers  
1. .311600  
/organism="Streptococcus pyogenes SSI-1"  
/mol\_type="genomic DNA"  
/strain="SSI-1"  
/db\_xref="taxon:193567"  
232..1587  
/gene="SP80001"  
232..1587  
/gene="SP80001"  
/note="similar to GB:AAF71535.1 (AF255728) percent  
identity 100 in 451 aa"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="chromosomal initiator protein"  
/protein\_id="BAC63096.1"  
/db\_xref="GI:28810158"  
/translation="MTENEQIFWNRVLEAQSOLKQATYEFVHDARLLKVDKHIATI

YLDQKELFWKLNKDVLTITAGFVYVNAQISVDYVFEEDLMIEQNQTKINQPKQOAL  
NSLPTVSDLSNYSFENIOGDRNRWAASIAVANTPTGTTITPLFIWGGGLGKTH  
LLNAGNSVLLENENARIKYITAEFNIFEVHILRLDMDLKEKFLVTRFKWGLTVNI  
SKAKTISGTOEEFNTFNHNNKQIVLTSRDTPDLNLEDLKFLVTRFKWGLTVNI  
TPPFETRVAILTNKIYENFIFPDQITTEYLAGOFDSNVRLEGALKDISLVANFKOI  
DTITVDIAABAIRAKODGPKMTVPIETEEIOAQVGFYVTKETKATKRTQNIVLAR  
QVAMFLAREMTDNSLPKIGKEFGGRDHSVTLHAYNKIKNWSQESLIEIETIKNKI  
K"  
1742..2878  
/gene="SP80002"  
1742..2878  
/gene="SP80002"  
/note="similar to GB:AAL96838.1 (AE009953) percent  
identity 99 in 378 aa"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="putative DNA polymerase III beta subunit"  
/protein\_id="BAC63097.1"  
/db\_xref="GI:28810159"  
/translation="MIQFSINRITLFIHALNATKRAISTKNAIPILSSIKIEVPTGVT  
LTSGNQISIENTIPVSNAGLLITSPGAILLEASFFINISLPDININVKTEQH  
QVLTSGKSEITLTKGVDQYPRLOEVSTENPLILTKLKSIIAETAPASLOSRRP  
ILTGHVILSNHKDFKAVATDSHMSORLITLNTSADFVDVVIKGSIRESAVTRDD  
IETVEFPSPQILFRSEHISFYTRLLGNYPTDRLMLTFEFTEVWFTQSLRAME  
RAFLISNATQNGTVKLEITQNHISAHVNSPEVKVNDLIDIVSQSGDLTISFNFTYL  
IESLKAIKSETQVTHFLSPVRPFTLTGDESESIQLITPVRTN"  
2953..3150  
/gene="SP80003"  
2953..3150  
/gene="SP80003"  
/note="similar to GB:AAK33148.1 (AE006472) percent  
identity 100 in 65 aa"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="BAC63098.1"  
/db\_xref="GI:28810160"  
/translation="MYQIGSFVEMKPKHACVIKGTGKANQKVLRVGADIKIOCTNC  
QHVIMSGRYDPERKLVQLP"  
3480..4595  
/gene="SP80004"  
3480..4595  
/gene="SP80004"  
/note="similar to GB:AAK33149.1 (AE006472) percent  
identity 98 in 371 aa"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="putative GTP-binding protein"  
/protein\_id="BAC63099.1"  
/db\_xref="GI:28810161"  
/translation="MALTAGIVGLPNVSKSTLENAITKAGAAANYPEATIDPNVGMV  
EVPDERLQKLTETLPKTVPTVETFTDIAGIVKASGEGELGNKFLANIEVDAIVH  
VVRAPDDNVNREGDAFVVDADITINLEILADLESINKRYAKRVARVATOK  
KESVAFNVLQIKPVLDEGKSARTIETDEAKVKGLFLITTKPVLVYRANVSDKV  
ANPDGIDYVKQIRDFAAENAEEVVIISARAEELLEDDEDEEFLAIGTESGVDK  
LTRAAYHLGLGTYFTAGEKVRATFKRGIKAPQACGIHSDFERGFIKAVTMSYDD  
LMTYSSEKAVKAGRLREGEKVVYVQDGDINEFENV"  
4665..5234  
/gene="SP80005"  
4665..5234  
/gene="SP80005"  
/note="similar to GB:AAK33150.1 (AE006472) percent  
identity 99 in 189 aa"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="putative peptidyl-tRNA hydrolase"  
/protein\_id="BAC63100.1"  
/db\_xref="GI:28810162"  
/translation="MVKMIIVGLGNPGSKYEKTKINIGPMAIDNIVKNLDVTTDDKNF



```
QY 1021 ATTAAGTGAACAGTAAATTCATTTCTTATCACAGTTCGACCATTCACCTTAACA 1080
Db 2762 ATTAAAGTGAACAGTAAATTCATTTCTTATCACAGTTCGACCATTCACCTTAACA 2821
QY 1081 CCAGCGATGAGGAGAAAGTTTATCCAAATTAATTAACACCAAGTACGACAAAC 1134
Db 2822 CCAGCGATGAGGAGAAAGTTTATCCAAATTAATTAACACCAAGTACGACAAAC 2875

RESULT 8
AR157860
LOCUS AR157860 3200 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 2 from patent US 6245906.
ACCESSION AR157860
VERSION AR157860.1 GI:16218875
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 3200)
Ueyama,H., Abe,K., Keshi,H. and Matsuhisa,A.
TITLES Probes for the diagnosis of infections caused by Streptococcus
pyogenes
JOURNAL Patent: US 6245906-A 2 12-JUN-2001;
FEATURES
source Location/Qualifiers
1..3200
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 74.9%; Score 849; DB 6; Length 3200;
Best Local Similarity 98.7%; Pred. No. 2.7e-146;
Matches 866; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 1 ATGATTCATTTTCAATTAATCGCATATTTATTCATGCTTTAAATACAACTAAACGT 60
Db 2324 ATGATTCATTTTCAATTAATCGCATATTTATTCATGCTTTAAATACAACTAAACGT 2383
QY 61 GCTATTAGCACTAAATGCGCATCTCTATCTTTCAATCAATAAATGAAGTCACTTCT 120
Db 2384 GCTATTAGCACTAAATGCGCATCTCTATCTTTCAATCAATAAATGAAGTCACTTCT 2443
QY 121 ACAGAGTAACTTTAACAGGGTCTAACGGTCAATATCAATTTGAAACACTATTCCTGTA 180
Db 2444 ACAGAGTAACTTTAACAGGGTCTAACGGTCAATATCAATTTGAAACACTATTCCTGTA 2503
QY 181 AGTAATGAAATGCTGGTTTGCTTAATTAATCTCTCAGGAGCTATTTATTAGAAGTAGT 240
Db 2504 AGTAATGAAATGCTGGTTTGCTTAATTAATCTCTCAGGAGCTATTTATTAGAAGTAGT 2563
QY 241 TTTTATTAATTAATTAATTTCAAGTTTGCAGATATTTAGTATATAATGTTTAAAGAAATGAA 300
Db 2564 TTTTATTAATTAATTAATTTCAAGTTTGCAGATATTTAGTATATAATGTTTAAAGAAATGAA 2623
QY 301 CAACACCAAGTTGTTTAAACCAAGTGTAAATCAGAGATTACCTTAAAGGAAAGATGTT 360
Db 2624 CAACACCAAGTTGTTTAAACCAAGTGTAAATCAGAGATTACCTTAAAGGAAAGATGTT 2683
QY 361 GACCAAGTATCTCGCTACAGAAGTATCAACAGAAATCCCTTGTATTTTAAACCAAAA 420
Db 2684 GACCAAGTATCTCGCTACAGAAGTATCAACAGAAATCCCTTGTATTTTAAACCAAAA 2743
QY 421 TTATTGAAGTCTATTATTGCTGAAAACAGCTTTTTCAGCAGCTTTTAAAGAAAGTCTGCT 480
Db 2744 TTATTGAAGTCTATTATTGCTGAAAACAGCTTTTTCAGCAGCTTTTAAAGAAAGTCTGCT 2803
QY 481 ATTTTAAACAGGATTCATATTGTTATTAAGTATATCAATAAGATTTTAAAGCAGTAGGACT 540
Db 2804 ATTTTAAACAGGATTCATATTGTTATTAAGTATATCAATAAGATTTTAAAGCAGTAGGACT 2863
QY 541 GACTCTCATCGTATGAGCCAAAGCTTTTAACTTTTGGAC- AATACTTTCAGCAGATTTGAT 599
Db 2864 GACTCTCATCGTATGAGCCAAAGCTTTTAACTTTTGGACAAATATCTTTCAGCAGATTTTGA 2923
```

```
QY 600 GGTAGTCTTCCAAATAAATCTTTGAGAGAATTTTTCAGCAGTATTTTACAGATGATTTGA 659
Db 2924 TGTGTTTATCCAAATAAATCTTTGAGAGAATTTTTCAGCAGTATTTTACAGATGATTTGA 2983
QY 660 GACCGTTCAGGATTTTCTCACCAGCCAAATCTTGTTCAGAAGTGAACACATTTCTTT 719
Db 2984 GACCGTTCAGGATTTTCTCACCAGCCAAATCTTGTTCAGAAGTGAACACATTTCTTT 3043
QY 720 TTATACACGCTCTTTAGAGGAAATTTATCCCGATACAGACCGCTTTTATTAATGACAGAA 779
Db 3044 TTATACACGCTCTTTAGAGGAAATTTATCCCGATACAGATCGTTTATTAATGACAGAA 3103
QY 780 TGAGACGAGGTTGTTTTCATATACCAATCCCTTCGCCACGCTATGGAACGTCCTTCT 839
Db 3104 TGAGACGAGGTTGTTTTCATATACCAATCCCTTCGCCACGCTATGGAACGTCCTTCT 3163
QY 840 GATTTCTAATGCTACTCAAAATGGTACTGTTAAAGCTT 876
Db 3164 GATTTCTAATGCTACTCAAAATGGTACTGTTAAAGCTT 3200

RESULT 9
E17380
LOCUS E17380 3200 bp DNA linear PAT 28-JUL-1999
DEFINITION gDNA derived from streptococcus pyogenes.
ACCESSION E17380
VERSION E17380.1 GI:5712063
KEYWORDS JP 1998262698-A/2.
SOURCE Streptococcus pyogenes
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 3200)
REFERENCE
1 (bases 1 to 3200)
AUTHORS Kamiyama,H., Fukuda,K., Karashi,H. and Matsuhisa,A.
TITLES DIAGNOSTIC PROBE FOR INFECTION DISEASE CAUSED BY STREPTOCOCCUS
PYOGENES MICROBE
JOURNAL Patent: JP 1998262698-A 2 06-OCT-1998;
COMMENT FUSO YAKUJIN KOGYO KK
OS Streptococcus pyogenes
PN JP 1998262698-A/2
PD 06-OCT-1998
PF 25-MAR-1997 JP 1997071077
PI KAWIYAMA HIROSHI, FUKUDA KANAKO, KARASHI HIROYUKI, PI
MATSUHIISA AKIO
PC C12Q1/68, C07H21/04//C12N15/09, (C12Q1/68, C12R1:46), (C12N15/09,
C12R1:46);
CC strandedness: Double;
CC topology: linear;
CC hypothetical: No;
FH Key Location/Qualifiers
FT source 1..3200
/strain="SP-7-44".
/organism="Streptococcus pyogenes"
location/Qualifiers
1..3200
/organism="Streptococcus pyogenes"
/mol_type="genomic DNA"
/db_xref="taxon:1314"

FEATURES
source
ORIGIN
Query Match 74.9%; Score 849; DB 6; Length 3200;
Best Local Similarity 98.7%; Pred. No. 2.7e-146;
Matches 866; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 1 ATGATTCATTTTCAATTAATCGCATATTTATTCATGCTTTAAATACAACTAAACGT 60
Db 2324 ATGATTCATTTTCAATTAATCGCATATTTATTCATGCTTTAAATACAACTAAACGT 2383
QY 61 GCTATTAGCACTAAATGCGCATCTCTATCTTTCAATCAATAAATGAAGTCACTTCT 120
Db 2384 GCTATTAGCACTAAATGCGCATCTCTATCTTTCAATCAATAAATGAAGTCACTTCT 120
Db 2384 GCTATTAGCACTAAATGCGCATCTCTATCTTTCAATCAATAAATGAAGTCACTTCT 2443
```

QY 121 ACAGGAGTAACTTTAAGAGGCTCTAACGAGTCAATGTCATATCAATTTGAAACACATATTCCTGTA 180  
 Db 2444 ACAGGAGTAACTTTAAGAGGCTCTAACGAGTCAATGTCATATCAATTTGAAACACATATTCCTGTA 2503  
 QY 181 AGTAATGAAATGCTGTTGCTAAATACCTCTCCAGAGCTATTTTATAGAGCTAGT 240  
 Db 2504 AGTAATGAAATGCTGTTGCTAAATACCTCTCCAGAGCTATTTTATAGAGCTAGT 2563  
 QY 241 TTTTATTAATATTAATTTCAAGTTTGCCAGATATTAGTATTAATGTTAAAGAAATGAA 300  
 Db 2564 TTTTATTAATATTAATTTCAAGTTTGCCAGATATTAGTATTAATGTTAAAGAAATGAA 2623  
 QY 301 CAACACCAAGTTGTTTAAACAGTGGTAATCAAGATTAACCTTTAAAGGAAAGATGTT 360  
 Db 2624 CAACACCAAGTTGTTTAAACAGTGGTAATCAAGATTAACCTTTAAAGGAAAGATGTT 2683  
 QY 361 GACCAATATCCTGCTCTACAGAGTATCAACAGAAATCCCTTTGATTTTAAACAAAA 420  
 Db 2684 GACCAATATCCTGCTCTACAGAGTATCAACAGAAATCCCTTTGATTTTAAACAAAA 2743  
 QY 421 TTATTGAAGTCTATTATTTGCTGAAACAGCTTTTGCCAGCAGTTTACAAGAAAGTCTGCT 480  
 Db 2744 TTATTGAAGTCTATTATTTGCTGAAACAGCTTTTGCCAGCAGTTTACAAGAAAGTCTGCT 2803  
 QY 481 ATTTTAAAGAGTTCATATTGTTATTAAGTAATCAATAAAGATTTTAAAGCAGTAGCGACT 540  
 Db 2804 ATTTTAAAGAGTTCATATTGTTATTAAGTAATCAATAAAGATTTTAAAGCAGTAGCGACT 2863  
 QY 541 GACTCTCATCGTATGAGCAAGCTTTAATCACTTTGGAC-ATACTTACGAGATTTGAT 599  
 Db 2864 GACTCTCATCGTATGAGCAAGCTTTAATCACTTTGGAC-ATACTTACGAGATTTGAT 2923  
 QY 600 GGTAGTCTTCCAGTAATCTTTGAGAGAAATTTTTCAGCAGATTTTACAGATGATTTGA 659  
 Db 2924 TGTGGTTATTCCAGTAATCTTTGAGAGAAATTTTTCAGCAGATTTTACAGATGATTTGA 2983  
 QY 660 GACCGTTAGGATTTTCTTCCAGCAAGCAATCTTTGTTTCAAGAGTGAACACATTTCTTT 719  
 Db 2984 GACCGTTAGGATTTTCTTCCAGCAAGCAATCTTTGTTTCAAGAGTGAACACATTTCTTT 3043  
 QY 720 TTATACACGCTCTTAGAGAGAAATTTATCCGATACAGACCGTTTATTAATGACAGAAAT 779  
 Db 3044 TTATACACGCTCTTAGAGAGAAATTTATCCGATACAGATCCTTTTATTAATGACAGAAAT 3103  
 QY 780 TGAGACGAGGTGTTTTCATACCAATCCCTTCGCCAGCTATGGAACGTCCTTCTT 839  
 Db 3104 TGAGACGAGGTGTTTTCATACCAATCCCTTCGCCAGCTATGGAACGTCCTTCTT 3163  
 QY 840 GATTCTTAACTGCTCAAAATGGTACTGTTAAGCTT 876  
 Db 3164 GATTCTTAACTGCTCAAAATGGTACTGTTAAGCTT 3200

RESULT 10  
 LOCUS AE014853 14925 bp DNA linear BCT 25-OCT-2002  
 DEFINITION Streptococcus mutans UA159 section 1 of 185 of the complete genome.  
 ACCESSION AE014853 AE014133  
 VERSION AE014853.1 GI:24376380  
 KEYWORDS  
 SOURCE Streptococcus mutans UA159  
 ORGANISM Streptococcus mutans UA159  
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 Streptococcus.  
 1 (bases 1 to 14925)  
 REFERENCE  
 AUTHORS Ajdic,D., McShan,W.M., McLaughlin,R.E., Savic,G., Chang,J.,  
 Carson,M.B., Primeaux,C., Tian,R., Kenton,S., Jia,H., Lin,S.,  
 Qian,Y., Li,S., Zhu,H., Najjar,F., Lai,H., White,J., Roe,B.A. and  
 Ferretti,J.J.  
 TITLE Genome sequence of Streptococcus mutans UA159, a cariogenic dental  
 pathogen  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (22), 14434-14439 (2002)

MEDLINE 22295063  
 PUBMED 12397186  
 REFERENCE 2 (bases 1 to 14925)  
 AUTHORS Ajdic,D., McShan,W.M., McLaughlin,R.E., Savic,G., Chang,J.,  
 Carson,M.B., Primeaux,C., Tian,R., Kenton,S., Jia,H., Lin,S.,  
 Qian,Y., Li,S., Zhu,H., Najjar,F., Lai,H., White,J., Roe,B.A. and  
 Ferretti,J.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-JUL-2002) Department of Microbiology and Immunology,  
 University of Oklahoma Health Sciences Center, 940 SL Young Blvd.,  
 Oklahoma City, OK 73104, USA  
 FEATURES  
 Location/Qualifiers  
 1..14925  
 /organism="Streptococcus mutans UA159"  
 /mol\_type="genomic DNA"  
 /strain="UA159"  
 /db\_xref="taxon:210007"  
 194..1552  
 /gene="dnaA"  
 /locus tag="SMU.01"  
 194..1552  
 /gene="dnaA"  
 /locus tag="SMU.01"  
 /note="Best Blastp Hit: gb|AAF71535.1|AF255728\_1  
 (AF255728) chromosomal initiator protein DnaA  
 [Streptococcus pyogenes]"  
 /codon start=1  
 /transl\_table=11  
 /product="chromosomal replication initiator protein, DnaA"  
 /protein\_id="AAN57794.1"  
 /db\_xref="GI:24376381"  
 /translation="MTENEQIFWNRVLEAQSQLQATDFVVSADAKLLKVEGNIAT  
 LLDMMKLFWKKNLQPVLTAGREVENTEISIEYVFETQSTNSPOISQNTAEALAT  
 ETLFPVQNDLNPYSPFNFIYIGENWAFASVSADLPGTITNFIYIGPGUGKTH  
 LLAIGNSVLASNPARIKIIISAENFVHRLQNMDELKFFNLDLLLDDIQ  
 SLAKSLAAATQEEFFFNALHNNKQIVLTSDRTDLDLEQLVTRFKWGLTVNI  
 TPDPETRAVLNNKIQYNSYSLSETIYLAGQFDSNVRDELKALXDISLVANFKKL  
 DVITVEAARAKODSPKMIPIEDIQKQVGFYGVTVKEIKSTRTONIVLA  
 ROVMYLAREMTDNLSPKIKFEGGRHSTVLHAYNKIKNLAQDSDLSKIEMETIKK  
 IK"  
 1708..2844  
 /gene="dnaN"  
 /locus tag="SMU.02"  
 1708..2844  
 /gene="dnaN"  
 /locus tag="SMU.02"  
 /EC\_number="2.7.7.7"  
 /note="Best Blastp Hit: sp|O06672|DP3B\_STRPN\_DNA  
 POLYMERASE III, BETA CHAIN >gi|2109446|gb|AAC45337.1|  
 (AF000658) beta subunit of DNA polymerase III  
 [Streptococcus pneumoniae]"  
 /codon start=1  
 /transl\_table=11  
 /product="putative DNA polymerase III, beta subunit"  
 /protein\_id="AAN57795.1"  
 /db\_xref="GI:24376382"  
 /translation="MIKFSINKYFFIQLALNATKRAISSKNAIPILSSLKTEVNSQST  
 LTSCGQISTENTISAEENAGLLVTSSGAILLEAFNFINVSSLPDITLDFEIEGH  
 OVLNKGSEITILKGDVEYQPLQSVGTNNPILITKLLKTIISETAFAARQESRP  
 ILGVLHLVNLNHEKFAVATDSHRMSQKLTLDHSSDDFDVVIIPSSLSRFAAVFTDD  
 IESVEFFNSQILFRSEYISFYRLLEGVPTDRLNGNFFETVVFNTNLYAME  
 RAHLISNATQNGTKLEIINNOVTAVNSPEVKNVEELDIESLGNLITISFNPTYL  
 IEALKAKSETVITRISPIRPTLTTPSDNSENFIQLITPVRTN"  
 3106..3297  
 /locus tag="SMU.05"  
 3106..3297  
 /locus tag="SMU.05"  
 /note="Best Blastp Hit: gb|AAC45338.1| (AF000658) OREX  
 [Streptococcus pneumoniae]"  
 /codon start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAN57796.1"



```
/db_xref="GI:24376383"
/translacion="MYSLGIVEMKKPHACTIKTTGGKANCWEVVRIGADIKIKCTN"
DHVVMGRHDPFRKIKYVL"
3461..4576
|locus_tag="SMU_06"
3461..4576
|locus_tag="SMU_06"
/note="Best Blastp Hit: dbj|BAB07770.1| (AP001520)
GTP-binding protein [Bacillus halodurans]"
/codon_start=1
/transl_table=11
/product="putative GTP-binding protein"
/protein_id="AAN57797.1"
/db_xref="GI:24376384"
/translacion="MALTAGIUGVPNTVKSTLFNAITKAGAAANYFPATIDPNVGVW
EVPDIRQLRELAVPKVTPTPEFTDIAGIVKGSGGLGNKPFIAREIDAIVH
VRAPNDENVREGRESDFVDPIADETINLELILADLESINKRVAERVKIAOTKO
KDSVAERNILOKIIPVELDGKSARTIOFTEDEAKIVKGFLLTKPVLYVANVEDOKV
ADPDSDVVQOIRFEAFANAENAVVISARAESEISELDDEDAEFLEAMGINSBGDK
LTRAAHYDLGLGYTFYAGEKEVRMATTKRGIKAPQAAGIIHSDFEKGTIFRAVTVSYDD
LIKYSKSAKVKAEGRLREEGDYIQVGDMIEPRFN"
4663..5232
|gene="pth"
|locus_tag="SMU_07"
4663..5232
|gene="pth"
|locus_tag="SMU_07"
/note="Best Blastp Hit: sp|P37470|PTH_BACSU PEPTIDYL-TRNA
HYDROLASE (PTH) (STAGE V SPOULATION PROTEIN C)
>gi|2127242|pir||S66083 stage v sporulation protein -
Bacillus subtilis >gi|7448588|pir||C69715 stage V
sporulation protein spovc - Bacillus subtilis
>gi|467442|dbj|BAAO5288.1|(D26185) stage V sporulation
[Bacillus subtilis] >gi|2632320|emb|CAB11829.1|(Z99104)
thermosensitive mutant blocks spore coat formation (stage
v sporulation) [Bacillus subtilis]"
/codon_start=1
/transl_table=11
/product="putative peptidyl-trna hydrolase"
/protein_id="AAN57798.1"
/db_xref="GI:24376385"
/translacion="MTKLIVLGNGPGSKYHETKNVGMALDRIVKPLNVQFTENTKTF
KAETVATFINGEKYFKPTPTFMNASGAFLAYNLTVEDLIYDLDMEVGKI
RFROQGASGHNGIKSIKIEGITQEPRIKIGIGRPKDKNVNYVLGKFDEKDKITI
FNTLDKDYKANYYLTQNKFEQTQKENG"
5234..8722
|gene="trcf"
|locus_tag="SMU_08"
5234..8722
|gene="trcf"
|locus_tag="SMU_08"
/note="Best Blastp Hit: gb|AAD03810.1| (AF054624)
transcription-repair coupling factor [Lactobacillus
sakai]"
/codon_start=1
/transl_table=11
/product="putative transcription-repair coupling factor"
/protein_id="AAN57799.1"
/db_xref="GI:24376386"
/translacion="MNILDFLKONQKI KEWTEHLSSAPTQLVMGFSSSKALAIAAAY
KQQKGKIVTSTQNEAKLATLSILIGEVEYVFADDVAAEFTFAQSDRTNSRV
ESLNFLQDKESGILLMSIVGIRLVSPONYQAQKLVKGYEYNLDNI IKOLSNIG
YQKVAOVNPCEFGSKRGIDLIVRNOPYPVRLFEGDEIDGIRTFDPDORSISLND
SFILI PSDDII L TSDSDFOACYNLEKALI I ANDNSLOPLYKELSVTDKHRYHDVRKF
LSFY PQEWELRIYPOTPULPTPNFHKGKLFYLYNFQYPMQPFNNOPPLLIDEIRRY
FYANNYKDWLRHQPATPTPNFHKGKLFYLYNFQYPMQPFNNOPPLLIDEIRRY
QKSEMVLTLQTSQSQRLOKLAGYFNLPISQVDNLKQAQALTIGNLSGTFYA
DEKVLVITEREI FHKKVCYKRRRSY ISNAERLKDYNELKGYVHVNHGFRFLGIE
TIYESVGHRYDLTYQONADRISI PIQIELLSKYVASDCGEPKINKINDGRFORIKQ
KVSQVEDIADLLALKYERSQLGFAFPSDDRKQOEFDNDPAYETEDQLRSIKEVK
KOMEDQMRLDVDGVGKTGVAMEARA KANDNDKQVALVPTTVLAQHVNTPKE
RFENFPVNAVMMSRFKTITQESETITTKLAGQDVIIIGTHRSSKQVTFPPDGLGVLD
EQORFGVKGHTKELKATKVDVLTATPTPTLHMMLGIRDLSVLETTPPSRPYPVO
```

TYVMETNASVIREAINREIDRGQIFYYINRVDTIDDKKVAELRELPEANIGFVHGK  
 SETQNLTLDDPISGSDYDLVVTTITIEDISNVNTLFENADYMLGSLTYQLRGV  
 GRSNLTAYAYLMYDPDKILTEVSEKLEAIKGFTELGSGFKIAMRDLISRGAGNIIGS  
 SOSGFTDSCFEMYSOLLBEATAAKQSGKRTSKGNAEINLOIDAVLPNTYISDERQK  
 IEIYKPIRESKTDYQALQDELIDRFQDYPDQVAYLIEIGLLKAYLDADFVSLDRK  
 QNSVVCPKESIKLKFYQDIFALSKTNLKAISENKGKIEIVDFVRAKQYEIIEE  
 LKFGGLKKLLKDKKK"  
 8852..9124  
 /locus\_tag="SMU.09"  
 8852..\_9124  
 /locus\_tag="SMU.09"  
 /notes="Best Blastp Hit: sp|P37557|YABO\_BACSU HYPOTHETICAL  
 9.7 KDA PROTEIN IN MFD-DIVIC INTERGENIC REGION  
 >gi|2127042|pir||S66089 conserved hypothetical protein  
 yabo - Bacillus subtilis >gi|467448|dbj|BAA05294.1|  
 (D26185) unknown [Bacillus subtilis]  
 >gi|2632326|emb|CAB11835.1|(Z99104) similar to  
 hypothetical proteins [Bacillus subtilis]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AA057800.1"  
 /db\_xref="GI:24376387"  
 /translation="MELDKYLKVSRIIKRSVAKEVSDKGRKIVNGILAKSSTDLLKIN  
 DTIEIRFGNKLJTVRVLEMKDSTGEDAVKVMVEIISETRIDLNGET"  
 9111..9482  
 /locus\_tag="SMU.10"  
 9111..\_9482  
 /locus\_tag="SMU.10"  
 /notes="Best Blastp Hit: gb|AAC98903.1|(AF023181) DivIC  
 homolog [Listeria monocytogenes]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AA057801.1"  
 /db\_xref="GI:24376388"  
 /translation="MEKPNIVQLNNKYINDEKTKRYEETKRRHRFIGWILFIIL  
 Query Match 60.9%; Score 690.8; DB 1; Length 14925;  
 Best local similarity 75.6%; Pred. No. 2.3e-117;  
 Matches 857; Conservative 0; Mismatches 277; Indels 0; Gaps 0;  
 Qy 1 ATGATTCAATTTTCAATTAAATCGCACATTATTTATTCATGCTTTTAAATACAACTAAACGT 60  
 Db 1708 ATGATAAAATTTTCAATTAAATAAAGTTTTTTTCTTACAGCCTTAAATGCTACCAAGCGA 1767  
 Qy 61 GCTATTAGCACTAAAAATGCCATTCCTATTCCTTTCATCAATAAAAAATGAAGTCACCTTCT 120  
 Db 1768 GCTATTAGTTCTAAAAATGCTATTCCTATTCCTTCTAGTTTAAAAAATGAAGTGAATTC 1827  
 Qy 121 ACAGGAGTAACTTTAAACAGGCTCTAAGCGTCAAAATCAATTCAGAAACACTATTTCCTGTGA 180  
 Db 1828 CAGTCAATTACTTTAACAGGCTCTAATGGACAAATTTCTATTGAAATACGATTTCACT 1887  
 Qy 181 AGTAATGAAATGCTGGTTTGCTAAATTACCTCTCCAGGAGCTATTTTATTAGAAGCTAGT 240  
 Db 1888 GAAGAAGAAATGCTGCACTATTGGTTACTTCTTCAGGAGCTATCTACTTGAAGCTAAT 1947  
 Qy 241 TTTTATTAATAATATATTTCAAGTTTGCAGATATTTAGTATATAATGTTTAAAGAAATTTGAA 300  
 Db 1948 TTCTTTTATTAATATTTGTTTCAAGCTTACCTGATATCACTTTAGATTTTGAAGAAATTTGAA 2007  
 Qy 301 CAACACCAAGTCTGTTTACCAGTGGTAAATCAGAGATTACCTTTAAAGGAAAAGATGTT 360  
 Db 2008 CAACATCAAGTCTGTTTAAATAGTGGAAATACGAAATTTACACTTAAAGGTAAGATGTG 2067  
 Qy 361 GAACAGTATCCTCGTCTCAAGAAGATATCAACAGAAAAATCCTTTTGATTTTTTAAAAACAAAA 420  
 Db 2068 GAACAATATCCTCGTTTGAGGAAGTTGGACAAATAATCTTTTAAATTTTAGAACAACAAA 2127  
 Qy 421 TTATTGAAGCTATTAATATGCTGAAACAGCTTTTTCAGCCAGCTTTTCAAGAAAGTCGTCCT 480  
 Db 2128 TTGTTAAAAACAATATATTTCTGAAACAGCTTTTTCGCTAGTACCAAGAAAGTCGACCA 2187



Query Match	56.5%	Score 641.2;	DB 6;	Length 1134;
Best Local Similarity	72.8%;	Pred. No. 5.1e-108;		
Matches 826;	Conservative	0;	Mismatches 308;	Indels 0;
				Gaps 0;

```
QY 1081 CCAGCGATGAGGAGAAAGTTTATCCAAATTAATTACACAGTACGACAAAC 1134
Db 1081 CCTGTGAGGATGACTGAAGATTTCATCAATTAATAACTCTCTGTCGTAAC 1134

RESULT 12
AX606582 1137 bp DNA linear PAT 17-FEB-2003
LOCUS Sequence 4511 from Patent WO02092818.
ACCESSION AX606582
VERSION AX606582.1 GI:28402110
KEYWORDS Streptococcus agalactiae
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
AUTHORS Glaser, P., Rueniok, C., Chevalier, F., Frangeul, L., Lalioui, L., and
Zouine, M., Couve, E., Buchrieser, C., Poyart, C., Trieu-Cuot, P., and
Kunst, F.
TITLE Streptococcus agalactiae genome sequence, use for developing
vaccines, diagnostic tools, and for identifying therapeutic targets
JOURNAL Patent: WO 02092818-A 4511 21-NOV-2002;
INSTITUT PASTEUR (FR); CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
source
1. .1137
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"

ORIGIN
Query Match 56.5%; Score 641.2; DB 6; Length 1137;
Best Local Similarity 72.8%; Pred. No. 5.1e-108;
Matches 826; Conservative 0; Mismatches 308; Indels 0; Gaps 0;

QY 1 ATGATTCATTTTCAATTAATCGCACATTATTATTTCATGCTTTTAAATACAACTAAAGCT 60
Db 1 ATGATTCATTTTCAATTAATTAATAAAATTTCTTTCGATGCTTAACGGTAAACCAACGA 60

QY 61 GCTATTAGCACTAAATGCCATTCCTATTCTTTCATCAATAAAATTTGAAGTCACTTCT 120
Db 61 GCTATTAGCATAAATGCGATTCCTCAATCTTTCATCTGTTTAAATAGAGTGAAGTAGG 120

QY 121 ACAGAGTAACTTTAACAGGGTCTAACCGTCAAATATCAATTTGAAAAACACTATTCTCTGA 180
Db 121 GATGCTATCAATTTAACCGGGTCAAATGGAACAAATTTCAATTTGAAAAACTATTCTCTGCT 180

QY 181 AGTAATGAAATGCTGTTTGTCTAATTAACCTCCAGGAGCTATTATTATTAGAGCTAGT 240
Db 181 TCAATGAAATGCGAGTTTACTAGTAACGAATCAGGCTCTATTTTGTTAGAAGCTGCT 240

QY 241 TTTTATTATTAATTAATTTCAAGTTTGCAGATATTAGTATAAATGTTTAAAGAAATTCGA 300
Db 241 TTTTATTATTAATTAATTTCAAGTTTACCAGATGTAATTTAGATATTACAGAGATTGAA 300

QY 301 CAACACCAAGTTGTTTAAACAGTGGTAAATCAGAGATTACCTTTAAAGGAAAAAGATGTT 360
Db 301 CAACATCAAAATGTTCTTACTAGTGGAAAAATCAGAGATTACTTTGAAAGGTAAGGATGTC 360

QY 361 GACCAATATCTCGTCTACAGAGATATCAACAGAAATCTTTGATTTTAAACAAAA 420
Db 361 GATCAATACCTCGTCTCAGGAAATGACACAGATATCTCCATTTAAACATTTAGAAGCTAAA 420

QY 421 TTATTGAAGTCTATTATTGCTGAAACAGCTTTTGCAGCGAGTTTACAAGAAAGTCTGCT 480
Db 421 CTGTTAAATCAATTAATTAAGAACTGCTTTTGTCTAGCCACACAGAAAGCGCTCA 480

QY 481 ATTTTAAACAGGAGTTTCATATTGTTATTAAGTAATCATTAAGATTTTAAAGCAGTAGCGACT 540
Db 481 ATCTTAAACAGGAGTTTCATTTGTTTATCAGTCAAAATAAATACTTTAAGGCTGTTGCGACA 540

QY 541 GACTCTCATGATGAGGCAACGGTTTAACTTACCTTTGGCAATACACTTTCAGCAGATTGATG 600
```

```
Db 541 GATTACACCCGTATGAGTCAACGCACACTTTCCAATTAGAGAAATCGGCTAAATTTTGAT 600
QY 601 GTAGTTCTTCCAAGTAAATCTTTGAGAGAAATTTTCAGCAGTATTTTACAGATGATATTGAG 660
Db 601 TTGGTTGTTCCAAGTAAATCCCTTCGAGAAATTTTCGGCTGTTTTTACAGATGATATTGAA 660

QY 661 ACCGTTGAGGTATTTTCTTCAACCAAGCAAAATCTGTTTCAGAGTGAACACATTTCTTTT 720
Db 661 ACTGTAGAGGTTTTCTTCTCAGATAGTCAAAATGTTTATTAGAAAGTGAATAATACAGTTC 720

QY 721 TATACACGCTCTTAGAAGAAATTTATCCGATACAGACCGGTTTATTATTAATGACAGAAAT 780
Db 721 TATACAGTTTGTCTTGAAGAAATCTACCTGATCTGATCGCTCTTAATCAATCAGTTT 780

QY 781 GAGACGGAGTTGTTTTCAATACCAATCCCTTCGCCACGCTATGGAACGTCCTTCTTTG 840
Db 781 GAAACCGAAATTAATCTTTTAAATACAAATGCTTTAGCCCATGCTATGGAACGTCCTTATTTA 840

QY 841 ATTTCTAATGCTACTCAAAATGGTACTGTTTAAGCTTGAATTTAGATTTACTCAAAATCATATTTCA 900
Db 841 ATTTGCAATGCAACTCAGAACGGTACTGTTTCGTTTGAAGAAATTTCAAAATGAGACAGTCTCA 900

QY 901 GCTCATGTTAACTCACCTGAGGTTGGTAAAGTAAACGAGGATTTAGATATTGTTAGTCTAG 960
Db 901 GCTCATGTTAACTCTCCAGAAAGTTGGTAAAGTAAATGAGGAATGGAATCTGTTAGTCTTT 960

QY 961 TCTGCTAGTATTAACTATCAGCTTCAATCCAATCTTCACTTATTTAGTCTTTTAAAGCT 1020
Db 961 AAAGTGATAGTTTAAATATTTAGCTTTTAAATCCAATCTTCACTTAAATTTGAATCTTTTAAAGCA 1020

QY 1021 ATTTAAAGTGAACAGTAAATTTCAATTTCTTATCACCAGTTCGACCATTCACCTTAACA 1080
Db 1021 GTAAAAAGCGAAACAGTTACGATTCGATTTATTTCTCCAGTACGTCCTTACTTTTGACA 1080

QY 1081 CCAGCGATGAGGAGAAAGTTTATCCAAATTAATTACACAGTACGACAAAC 1134
Db 1081 CCTGTGAGGATGACTGAAGATTTCATCCAAATTAATACTCTCTGTTGCTGACTAAC 1134

RESULT 13
AX6014191 21262 bp DNA linear BCT 23-SEP-2002
LOCUS Streptococcus agalactiae 2603V/R section 1 of 100 of the complete
genome.
ACCESSION AE014191 AE009948
VERSION AE014191.1 GI:22533002
KEYWORDS Streptococcus agalactiae 2603V/R
SOURCE Streptococcus agalactiae 2603V/R
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 21262)
Tetelin, H., Maignani, V., Cieslewicz, M.J., Eisen, J.A.,
Peterson, S., Wessels, M.R., Paulsen, I.T., Nelson, K.E., Margarit, I.,
Read, T.D., Madoff, L.C., Wolf, A.M., Beanan, M.J., Brinkac, L.M.,
Dougherty, S.C., DeBoy, R.T., Durkin, S., Kolonay, J.F., Umayam, L.A.,
Madupu, R., Lewis, M.R., Radune, D., Fedorova, N.B., Scanlan, D.,
Khouri, H., Mulligan, S., Carty, H.A., Cline, R.T., Gill, J.,
Scarselli, M., Mora, M., Iacobini, E.T., Brettoni, C., Galli, G.,
Mariani, M., Vegni, F., Maione, D., Rinaudo, D., Rappuoli, R.,
Telford, J.L., Kasper, D.B., Grandi, G. and Fraser, C.M.
Complete genome sequence and comparative genomic analysis of an
emerging human pathogen, serotype V Streptococcus agalactiae
Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12391-12396 (2002)
12200547

2 (bases 1 to 21262)
Tetelin, H., Maignani, V., Cieslewicz, M.J., Eisen, J.A.,
Peterson, S., Wessels, M.R., Paulsen, I.T., Nelson, K.E., Margarit, I.,
Read, T.D., Madoff, L.C., Wolf, A.M., Beanan, M.J., Brinkac, L.M.,
Dougherty, S.C., DeBoy, R.T., Durkin, S., Kolonay, J.F., Umayam, L.A.,
Madupu, R., Lewis, M.R., Radune, D., Fedorova, N.B., Scanlan, D.,
Khouri, H., Mulligan, S., Carty, H.A., Cline, R.T., Gill, J.,
```

Scarselli, M., Mora, M., Iacobini, E.T., Brettoni, C., Galli, G.,  
Mariani, M., Vegni, F., Malone, D., Rinaudo, D., Rappuoli, R.,  
Telford, J.L., Kasper, D.L., Grandi, G. and Fraser, C.M.  
Direct Submission  
JOURNAL  
Submitted (18-JUL-2002) The Institute for Genomic Research, 9712  
Medical Center Dr. Rockville, MD 20850, USA  
FEATURES  
Location/Qualifiers  
Source  
1. .21262  
/organism="Streptococcus agalactiae 2603V/R"  
/mol\_type="genomic DNA"  
/strain="2603V/R"  
/serotype="v"  
/db\_xref="taxon:208435"  
102. .1463  
/gene="dnaA"  
/note="synonym: SAG0001"  
102. .1463  
/gene="dnaA"  
/note="identified by match to PFAM protein family HMM  
PF00308"  
/codon\_start=1  
/transl\_table=11  
/product="chromosomal replication initiator protein DnaA"  
/protein\_id="AAM98909.1"  
/db\_xref="GI:22533003"  
/translation="MTENEQLFWRVRLSRSQIAAPAYEFFVLEARLLKIEHOTAVI  
TLDNTEMKLFWQNLGVILTAGFEIIFNAEITANYVNSDLHLOETSPSNYQOSNEV  
NLPFRKIDSLKERYTFANFVQGDENWASIAVADSPGTTNPLFIWGGPLGLG  
THLLAIGNVLDRNPANRVLVITAENFNEFVSIRLDSMEELKEKERNLDLLID  
IQSLAKTKLGGTQEFTFNALHNDQIVLTSDRNPQNLDEERLVTRPSWGLPV  
NITPDPFTRVALLTNKIQEYPDPQDTIEYLFAGEFDSNVRELEGALKNISLVADF  
HAKTIVDIAEAIKARNKDGPIVTVPIEIQIQVGFYGVTVKEIKATKTDQIVL  
ARQVAMYAREMNTDNLSPKIGKEFGGRDHTVLHAYNKIKNNVAQDDNLRLEIETIKN  
KIR"  
1618. .2754  
/gene="dnaN"  
/note="synonym: SAG0002"  
1618. .2754  
/gene="dnaN"  
/note="identified by match to PFAM protein family HMM  
PF00712"  
/codon\_start=1  
/transl\_table=11  
/product="DNA polymerase III, beta subunit"  
/protein\_id="AAM98910.1"  
/db\_xref="GI:22533004"  
/translation="WIHFSINKNPFHLATVTKRAISHKNAIPILSTVKIEVTRDAII  
LTGNGQISIENTIPASNENAGLVNPGSILLRAGPFINIISLPDVTLEFTEIOH  
QIVLTSGKSEITLKGKVDVQYPRLEQMTDTPLILEYKLLASINETAPASQOESRP  
ILTVGLVISONKFKAVADSHRMSQRTFOLEKSNFIDLVPVPSKSLRPSAVFTDD  
IETVEVPSDSMLFRSENISFYTLRLEGNYPDTRLTNOFETEIIIFNTNALRHAME  
RAYLISNATQNGTVRLREIQNETVSAHNSPEVGNBELDVTSLKGSLSNISFNPTYL  
IESLKAVKSETVTRIFISVPVFPFTLTGDEDTFQILTTPVRTN"  
2824. .3705  
/gene="SAG0003"  
2824. .3705  
/gene="SAG0003"  
/note="identified by match to PFAM protein family HMM  
PF00781"  
/codon\_start=1  
/transl\_table=11  
/product="diacylglycerol kinase catalytic domain protein,  
putative"  
/protein\_id="AAM98911.1"  
/db\_xref="GI:22533005"  
/translation="MTLYLIANPHAGNKNASTIVGKIOELVHTEDISVFYTEOKDKK  
KQVINILSPKESDHLMTIGGDGLTSKVMYLPQHIPTVYPVPGSGNDPARKLPGH  
KETLTATQELRCFVFDKGLILNSLDLGFAYVWVWASKNIKNINLRKGLNI  
TYIVIAKSLHSSKQVULVEGTGQQIKLNDLYFFALANNYFGGGITITWPKASALT  
AELDMVAKGHTLURKLSILLSLVFKRHTTSKSKTHQTFKAMTVYFPKNSLIIDGSI  
VELDQISLKQCKRYLM"  
3715. .3912  
/gene="SAG0004"

3715. .3912  
/gene="SAG0004"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAM98912.1"  
/db\_xref="GI:22533006"  
/translation="MYQVGSVLVENKKPHACVIKETGKANQWVLRVGADIKIQCTNC  
QHVIMSRYDFERKLLKVLQP"  
4114. .4317  
/gene="SAG0005"  
4114. .4317  
/gene="SAG0005"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAM98913.1"  
/db\_xref="GI:22533007"  
/translation="MDNLIRLAEIFDVSLEHVLAKTNEVKVERIYENKPLDLRKYNK  
IYWFIPNIRSLIILAILLFL"  
4477. .5592  
/gene="ychF"  
/note="synonym: SAG0006"  
4477. .5592  
/gene="ychF"  
/note="identified by match to TIGR protein family HMM  
TIGR00092"  
/codon\_start=1  
/transl\_table=11  
/product="GTP-binding protein YchF"  
/protein\_id="AAM98914.1"  
/db\_xref="GI:22533008"  
/translation="MALTAGIVLGNVSKTLFNAITKAGAEAAANYFPATIDPNVGMV  
EVPDERLQKLELITPKKTVPTTFDITDIAGIVKGSKGGLGNKFLANIREVDVIVH  
VVRADFDDENVMREQREDADFVADIADINLELILADLESINKRYARVEKMAITQD  
KESVAEPNVLOKIPVLEDCGKARTTFTPEEAKVVKGLFLTTTPVLYVANVEDVK  
ADPDIDYVQIRAFETENAEVVIISARAEELISELDDKLEFLFALIGLTSQGVK  
LTRAAYHLGLTYFTAGEKEVRWTFKGIKAPQAASIIHSDFERGFIRAVTMSYDD  
LIQVSEKAVKEAGRLKEGKEYIVQGDIMEPRFNV"  
5676. .6251  
/gene="pth"  
/note="synonym: SAG0007"  
5676. .6251  
/gene="pth"  
/note="identified by match to PFAM protein family HMM  
PF01195"  
/codon\_start=1  
/transl\_table=11  
/product="peptidyl-tRNA hydrolase"  
/protein\_id="AAM98915.1"  
/db\_xref="GI:22533009"  
/translation="MVNMIVGLGNPGSKNDTKHNGFMADVIRVDKLDVNFTEKRF  
KAEISDFINGEKIYFIKPTFTNNSGIAVKALLTYNTSIKDMIIYDDLDMVGKI  
RFRQKSAGGHNGIKSIIHLGTQEFDRIVKVGPRNRMVTINHVILGKFKDNDIMI  
LNTLDKVDNAVNYLQTNDFQKTMQYNGLK"  
6248. .9745  
/gene="mfd"  
/note="synonym: SAG0008"  
6248. .9745  
/gene="mfd"  
/note="identified by match to PFAM protein family HMM  
PF00270"  
/codon\_start=1  
/transl\_table=11  
/product="transcription-repair coupling factor"  
/protein\_id="AAM98916.1"  
/db\_xref="GI:22533010"  
/translation="WNIELFSONKVVRVTHSGVTNSRQLVMGFSKASKAIAIASAY  
EKLKSKIMVVTATQTDSDKLSLSDISSLIGEDNVYQFADDDVPAAEIIFSLDKISRL  
SALRFLKDPKNGVLITSISGLRLLLPNEFVSKSQYKEIGQECVLDKCKNLVNLG  
YKQSVQVFSFGFSQRGDINDIFEMTQETPYRLEFPGDEIDGIRQDFDIDTQKSLKQLE

SVLSIPDADLILQADOPFERAKKLEGVLYNTASEVQRTYLSSEVLSTSTENHPFKSDIRRI  
LSIYPKEWIGLILYIPIEGTPTLFDVDDFQKIVDRNAKDLLEIASLTLESDQLQCKSHSSIN  
YFSDPTQKQLQYQATIFFSNFHKLGNLKFEDKLHHFTYQGMQEFFNQFPULLVDENIRY  
KKSGATVLLQVDSKGNLLQENLKYGLDLIIISDKNDIVOKESQLIVGLHNSGFYPA  
DEKIVLITEREYTHRVKRLKIRRSNINAEKLDIYNELSVGDTYVHNHVHGVSGFKLGE  
TIBIQGHRDIADLLTLQYQADNRIISIPVEQIBLLTKYVSADCKEPIKTNLNDGRFKKAQ  
RVAQOVEDIADLLTKYAESRQLQGFAPSDPDNMQNDFNDAFVTEQDLRSIKIK  
QDMENRPMRLLRVSDRGCKTEVAMRAAPKAVNDHKQVVVLVPTTLAQOHPENFKE  
RFSNPYTVDLVSRPSKKEQTTDLKRLSKQVDIIIGTHRLLSQDVPSLDGLLVID  
EQRGVGVKHEKELKELKTKVDVLTLTATPIPRILHMSMLGIRLSVIEPTPTNRYPVQ  
TYVLENTPLIRRAIRIREDRGQVFVITGVNDVIIDQKVELQELPVEASIGYVHGVM  
SEIOLENTLIDINGDYVDVATIIETGVNDIVNANTLPVENADHMLSTYLYQLGRG  
GSGNRATAYAYMPKPTKTEISERKDLAIKGFTELGSFKIAMRDLSTRAGNAILGA  
QSNRTDVSDFEMYSQLLEQAIATKQCKSLIRQKGAELQAIDYDLPAEYISDERQK  
IETKIRIRELRETRADYEAQOELIDRFEGYDPQVAYLLEIGLLKAYLDLAFTELVBK  
GNBISILFEKASLKYFLTYQDFEALSQTQLKARISETNGWMEVVFPNIKHKONYEIIIE  
LKFACFCFIEIKSRKPEE"

9906. .10001  
/gene="SAG0009"  
9906. .10001  
/gene="SAG0009"  
/note="identified by Glimmer2; putative"  
/codon\_start=1

Query Match	56.5%	Score 641.2	DB 1	Length 21262
Best Local Similarity	72.8%	Pred. No. 2.8e-108		
Matches 826	Conservative 0	Mismatches 308	Indels 0	Gaps 0
Qy	1	ATGATTCAAATTTCAATTAATCGCACATTAATTTATTCATCGTTTAAATCAACTAACGTT	60	
Db	1618	ATGATTCAATTTTCAATAAATAAAAAATTTCTTCTGTGATGCTCTAAACGGTAAACCAACGA	1677	
Qy	61	GCTATTAGCACTAAAAATGCAATTCCTTAATTCCTTTCATCAATAAAAAATGGAAGTCACCTTCT	120	
Db	1678	GCTATTAGTCATAAAAATGCGATCCAAATCCTTTCACCTGTTAAATAGAAAGTCACCTAGG	1737	
Qy	121	ACAGGAGTAATCTTTAACAGGGTCTAACGGTCAAAATATCAAAATGAAACACATATTCCTGTGA	180	
Db	1738	GATGCTATCAATTTTAAACGGGGTCAAAATGGACAATAATTTCAATTCGAAAATACATATTCCTGCT	1797	
Qy	181	AGTAATCAAAATGCTGGTGTGCTAAATTAACCTCTCCAGGAGCTATTTTATTAGAAAGCTAGT	240	
Db	1798	TCAAATGAAAATGCGAGTGTACTAGTAACGAATCCAGGCTCTATTTTGTGTAGAAAGCTGGT	1857	
Qy	241	TTTTTTTAAATATTTATTTCAAAGTTTGCCAGATATTAGTATAATGTGTAAGAAATTTGAA	300	
Db	1858	TTCTTTATTAATATTTATTTCAAGTTTACAGATGTAATCTTTAGAAATTTACAGAGATTGAA	1917	
Qy	301	CAACACCAAGTTGTTTTAACACAGTGGTAAATTCAGAGATTACCTTTAAAGGAAAGATGTT	360	
Db	1918	CAACATCAAAATTTGTTCTTACTAGTGGAAATTCAGAGATTACTTTTGAAGGTGAAGGATGTC	1977	
Qy	361	GACCAGTATCCTGCTACAGAAGTATCAACAGAAATCCTTTGATTTTAAAAAACAAA	420	
Db	1978	GATCAATACCCCTCGTCTGCAGGAAATGCAACACAGATATCTCCATTAACATTTAGAAACTAAA	2037	
Qy	421	TTATTGAAGTCTATTATTTGCTTGAAACAGCTTTTTCAGCCAGTTTACAAAGAAAGTCGTCCT	480	
Db	2038	CTGTTAAAAATCAATTTAATGAACATGCTTTTGTGCTGACCCAAACGAAGAACCCGTTCCA	2097	
Qy	481	ATTTTAAACAGGAGTTCAATTTGTTATTAAAGTAATCATAAAGATTTTAAAGCAGTAGCGACT	540	
Db	2098	ATCTTAAACAGTGTTTCAATTTGGTTATTCAGTCAAAAATAAATACITTTAAGGCTGTTGGACA	2157	
Qy	541	GACTCTCATCGTATGAGCCAAAGTTTAAATCACTTTGGACAATTAATCTTCAGCAGATTTGATG	600	
Db	2158	GATTTCACACCGTATGAGTCAACCGCACCTTTTCCAAATTTAGAGAAATCGGCTAATAATTTTGAT	2217	
Qy	601	GTAGTCTTCCCAAGTAAATCTTTTGAGAGAAATTTTTCAGCAGTATTTTACAGATGATATTGAG	660	
Db	2218	TTGGTTGTTTCAAGTAAATCCCTTCGAGAAATTTTTCGGCTGTTTTCAGAGATGATATTGAA	2277	
Qy	661	ACGTTTGAGGTATTTTCTCACCAAGGCCAAATCTTGTTCAGAAAGTGAACATATTTCTTTT	720	

Db	2278	ACTGTAGAGGTTCTTCTCAGATAGTCAAAATGTTATTTAGAAAGTGAATAATACGCTTC	2337
Qy	721	TATACACGCCCTCTTTAGAAAGAAATATCCCGATACAGACCGTTTATTAATATGACAGAAATTT	780
Db	2338	TATACACGTTTGCTTGAAGGAACTACCCGTGATCTGATCGCCTCTTAATAATCAGTTT	2397
Qy	781	GAGACGAGGTTGTTTCAATACCCAATCCCTTCGCGCACGCTATGGAACGTGCTTCTTGG	840
Db	2398	GAACCCGAAATATCTTTAATACAAATGCTTTACGGCATGCTATGGAACGTGCTTATTTA	2457
Qy	841	ATTTCCTAATCTACTCAAAATGCTACTGTTAAAGCTTTGAGATTACTCAAAATCATATTTCA	900
Db	2458	ATTTCGAATGCAACTCAGAACGGTACTGTTTCGTTTAGAAATTCAAAATGAGACAGTCTCA	2517
Qy	901	GCTCATGTTAACTCACCTGAGGTTGGTAAAGTAAACGAGGATTTAGATATTTGTTAGTCAG	960
Db	2518	GCTCATGTTAACTCTCCAGAAAGTTGGTAAAGTTAAAGGAAATGAGTACTGTTAGTCTT	2577
Qy	961	TCGTGGTAGTTAACTATCAGCTTCAATCCAACTTACCTTTATGAGTCTTTTAAAGCT	1020
Db	2578	AAAGGTGATAGTTTAAATATTTAGCTTTTAAATCCAACTTACCTTAATGAACTTTTAAAGCA	2637
Qy	1021	ATTTAAAGTCAACAGTAAAAATTCATTTCTATCACCAGTTTCGACATTCACCTTAACA	1080
Db	2638	GTAAGGACGAAACAGTACGATTTCGATTATTTCTCCAGTACGTCATTTACTTTTGACA	2697
Qy	1081	CCAGGCGATCAGGAAGAAGTTTTATCCAAATTAATACACCAGTACGAAACAAAC	1134
Db	2698	CCTGGTGAAGTACTGAAGATTTTCATCCAAATTAATTAATCTCTGTCGTACTAAC	2751
RESULT 14			
AX602184			
LOCUS	AX602184	26144 bp	DNA linear PAT 17-FEB-2003
DEFINITION	Sequence 113 from Patent WO02092818.		
ACCESSION	AX602184		
VERSION	AX602184.1	G1:28402053	
KEYWORDS			
SOURCE	Streptococcus agalactiae		
ORGANISM	Streptococcus agalactiae		
REFERENCE	1		
AUTHORS	Glaser,P., Rusniok,C., Chevalier,F., Frangeul,L., Lalioui,L., Zouine,M., Couve,E., Buchrieser,C., Poyart,C., Trieu-Cuot,P. and Kunst,F.		
TITLE	Streptococcus agalactiae genome sequence, use for developing vaccines, diagnostic tools, and for identifying therapeutic targets		
JOURNAL	Patent: WO 02092818-A 113 21-NOV-2002; INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)		
FEATURES	Location/Qualifiers		
source	1. .26144		
	/organism="Streptococcus agalactiae"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:1311"		
ORIGIN			
Query Match	56.5%	Score 641.2;	DB 6; Length 26144;
Best Local Similarity	72.8%	Pred. No. 2.7e-108;	
Matches	826; Conservative	0; Mismatches 308; Indels	0; Gaps 0;
Qy	1	ATGATTCATATTTCAATTAATCCGACATATTTATTCATGCTTTAAATACACTAAACGT	60
Db	10096	ATGATTCATATTTCAATTAATTAATTAATTTCTTCTGATGCTCTAACGGTAAACCAACGA	10151
Qy	61	GCTATTAGCACTAAAAATGCGATTCCTATTCTTTTCATCAATAAAAAATTGAAGTCACCTCT	120
Db	10156	GCTATTAGTCATAAAAATGCGATTCATCTTTCAATCTTTCACTGTTTAAATAGAGTACTAGG	10211
Qy	121	ACAGGAGTAACTTTTAAACGGGCTCTAACGGTCAAAATATCAATTGAAAAACACTATTCCTGTGA	180

Db 10216 GATGCTATCAATTTTAAACGGGCTCAATGGACAAATTTCAATTTGAAATATCTATTCCTGCT 10275  
 Qy 181 AGTAATGAATAGTGTGTTGCTAATTAACCTCTCCAGGAGCTATTTATTAGAGCTAGT 240  
 Db 10276 TCAATGAATAGTGTGTTGCTAATTAACCTCTCCAGGAGCTATTTATTAGAGCTAGT 10335  
 Qy 241 TTTTATTATTAATATTTTCAAGTTTCCAGATATTTAGTAAATGTTTAAAGAAATTTGAA 300  
 Db 10336 TTTTATTATTAATATTTTCAAGTTTCCAGATATTTAGTAAATGTTTAAAGAAATTTGAA 10395  
 Qy 301 CAACACCAAGTGTGTTTAAACAGTGTGTAATACAGATATTTACCTTAAAGAAAGATGTT 360  
 Db 10396 CAACATCAAAATGTTTCTTACTAGTGGAAATCAGATATTTTGAAGGTAGGATGTC 10455  
 Qy 361 GACCAATATCTGCTCTACAGAAAGTATCAACAGAAATCTTTGATTTTAAACAAAA 420  
 Db 10456 GATCAATATCTGCTCTGACAGAAATGACAAACAGATCTCCATTAATTTAGAACTAAA 10515  
 Qy 421 TTATTGAAGTCTATTTATGCTGAAACAGCTTTTGCAGCCAGTTTACAAGAAAGTCTGCTCT 480  
 Db 10516 CTGTTAAATTCATATTTAATGAACCTGCTTTTGTCTAGCCAAACAGAAAGCCGTCCA 10575  
 Qy 481 ATTTTAAACAGAGTCTCATATTTGTTAATGATTAATCAATAAAGATTTTAAACAGTAGGACT 540  
 Db 10576 ATCTTAAACAGTGTCTCATTTGTTTATCAGTCAAAATAAATATCTTTAAGGCTGTTGCGACA 10635  
 Qy 541 GACTCTCATGTAAGAGCAACAGTTTAAATCACTTTGGACAATATCTTACAGATTTGATG 600  
 Db 10636 GATTCACAGCTATGATGATCAACGACCTTTTCCAATTTAGAGAAATCGGCTAATTTTGAAT 10695  
 Qy 601 GTAGTTCTTCCAGTAATCTTTGAGAAATTTTACAGAGTATTTTACAGATGATATGAG 660  
 Db 10696 TTGTTGTTTCCAGTAATCTTCCCTTGAGAAATTTTCCGCTGTTTATACAGATGATATGAA 10755  
 Qy 661 ACCGTTGAGGATTTTCTTCCACCAAGCAAACTCTTGTTCAGAAAGTGAACACATTTCTTTT 720  
 Db 10756 ACTGTAGAGGTTTCTTCTCAGATAGTCAAACTTTTATTTAGAGTGAATATCAGCTTC 10815  
 Qy 721 TATACAGCTCTTACAGAAATTTATCCGATACAGACCGTTTATTAATGACAGAAATTT 780  
 Db 10816 TATACAGCTTCTGTTGAAGAAATCTACCTGTATGATGATGCTCTTAACTAATCAGTTT 10875  
 Qy 781 GAGACGAGGTTGTTTCTCAATACCAATCCCTTCCGACGCTATGGAACGCTCTTCTTG 840  
 Db 10876 GAAACCGAAATTTATCTTTTAAACAAATGCTTTACGCCATGCTATGGAACGCTTATTTA 10935  
 Qy 841 ATTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTGAGATTTACTCAAAATCATATTTCA 900  
 Db 10936 ATTTCCGATGCAACTCAGAACGCTACTGTTGTTTGAATTTCAAAATGACACAGTCTCA 10995  
 Qy 901 GCTCATGTTAACTCACTGAGTTGGTAAAGTTAAAGTAAATGAGGAAATGGAATCTGTTAGTCTT 11055  
 Db 10996 GCTCATGTTAACTCTCCAGAGTTGGTAAAGTTAAAGTAAATGAGGAAATGGAATCTGTTAGTCTT 11055  
 Qy 961 TCTGTAGTGAATTTAACTPATCAGCTTCAATCCAACTTACCTTTATGAGTCTTTTAAAGCT 1020  
 Db 11056 AAAGGTAGTGTAAATATTTAGCTTTAATCCAACTTACCTAAATGATCTTTTAAAGCA 11115  
 Qy 1021 ATTTAAAGTGAACAGTAAATTTCAATTTCTATCAACAGTTTCAACCTTACCTTAAAC 1080  
 Db 11116 GTAAACCGAACAGTATGCAATGATTTATTTCTCCAGTACGCTTACCTTACTTTGACA 11175  
 Qy 1081 CCAGCGATAGGAAAGTGTATTTTATCCAAATTAATTTACACAGTACCAACAAAC 1134  
 Db 11176 CCTGTGAAGATCTGAAGATTTTCAATCCAAATTAATTTACTCTGTTCTGCTACTAAC 11229

RESULT 15  
 CQ655074  
 LOCUS CQ655074 60266 bp DNA linear PAT 02-FEB-2004  
 DEFINITION Sequence 12031 from Patent WO234711.  
 ACCESSION CQ655074  
 VERSION CQ655074.1 GI:41687951

KEYWORDS  
 SOURCE  
 ORGANISM

Streptococcus agalactiae  
 Streptococcus agalactiae  
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 Streptococcus.

REFERENCE

1 Telford, J., Masignani, V., Margarit y Ros, I., Grandi, G., Fraser, C.  
 and Tettelin, H.  
 Nucleic acids and proteins from streptococcus groups a & b  
 Patent: WO 0234771-A 12031 02-MAY-2002;  
 Chiron S.p.A. (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)  
 Location/Qualifiers

FEATURES  
 source

1..60266  
 /organism="Streptococcus agalactiae"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:1311"  
 /note="seq 10967 to big: splitted in-seq 10.967, from  
 0.000.001 to 0.349.980-seq 12.025, from 0.300.001 to  
 0.649.980-seq 12.026, from 0.600.001 to 0.949.980-seq  
 12.027, from 0.900.001 to 1.249.980-seq 12.028, from  
 1.200.001 to 1.549.980-seq 12.029, from 1.500.001 to  
 1.849.980-seq 12.030, from 1.800.001 to 2.149.980-seq  
 12.031, from 2.100.001 to 2.160.266"

ORIGIN

Query Match 56.5%; Score 641.2; DB 6; Length 60266;  
 Best Local Similarity 72.8%; Pred. No. 2.2e-108;  
 Matches 826; Conservative 0; Mismatches 308; Indels 0; Gaps 0;

Qy 1 ATGATTCAAATTTTCAATTAATCGCACATTTATTTATTCATGCTTTTAAATACAACTAAACGT 60  
 Db 33097 ATGATTCATTTTCAATTAATTAATTTCTTTCTGATGCTCTAACGGTAAACCAACGA 33156  
 Qy 61 GCTATTAGCACTAAAATGCCATTTCTTATTTCTTCATCAATAAAAATTTGAAGTCACCTTCT 120  
 Db 33157 GCTATTAGTCATAAAAATGCCATTTCCAATCTTTCAACTCTTAAATAGAAAGTACCTAGG 33216  
 Qy 121 ACAGGAGTAACTTTAAACAGGCTCTAAGGCTCAATATCAATTCATGAAACACATTTCTCTGA 180  
 Db 33217 GATGCTATCATTTTAAACGGGCTCAATGGAACAAATTTCAATTTGAAATATCTATTTCTGCT 33276  
 Qy 181 AGTAATGAAATGCTGTTTGTCTTAAATTAATCTCTCCAGGAGCTATTTTATTAGAAGTGTAGT 240  
 Db 33277 TCAATGAAATGCTGTTTACTAGTAAACGAATCCAGGCTCTATTTTGTAGAAGTGTGT 33336  
 Qy 241 TTTTATTAAATATTTTCAAGTTTCCAGATATTAGTATAAATGTTTAAAGAAATTTGAA 300  
 Db 33337 TTTTATTAAATATTTTCAAGTTTACCAGATGTTAACTTTAGAATTTTACAGAGATTTGAA 33396  
 Qy 301 CAACACCAAGTGTGTTTAAACAGTGGTAAATCAGAGATTTACCTTAAAGGAAAGATGTT 360  
 Db 33397 CAACATCAAAATTTCTTACTAGTGGAAATCAGAGATTTACTTTGAAGGTAGGATGTC 33456  
 Qy 361 GACCAATCTCTGCTCTACAGAGATTCACAGAAATCTTTGATTTTAAACAAACAAA 420  
 Db 33457 GATCAATACCTCTGCTGAGGAAATGACACAGATCTCTCCATTAACATTAGAAACTAAA 33516  
 Qy 421 TTATTGAAGTCTATTATTTGCTGAAACAGCTTTTGCAGCCAGCTTTCAAGAAAGTCTGCTCT 480  
 Db 33517 CTGTTAAATCAATTTAATGAACTGTTTGTCTGCTAGCCACAGAAAGCCGTCCA 33576  
 Qy 481 ATTTTAAAGGAGTTCATATTTGTTAATGATTAATCAATGATTTTAAAGCAGTAGGCTCT 540  
 Db 33577 ATCTTAAAGGAGTTCATTTGTTTATCAGTCAAAATTAATATCTTTAAGGCTGTTGCGACA 33636  
 Qy 541 GACTCTCATGCTATGAGCCAAAGCTTTTATCACTTTTGGCAATCTTTCAGCAATTTGATG 600  
 Db 33637 GATTACACCGTATGAGTCAACGCACTTTCCAATTAGAGAAATCGGCTAATTAATTTGAT 33696  
 Qy 601 GTAGTTCTTCCAGTAAATTTTGTGAGAGAAATTTTTCAGCAGTATTTTACAGATGATATTTAG 660  
 Db 33697 TTGGTTGTTCCAAAGTAAATCCCTTCAGAAATTTTTCGGCTGTTTTCACAGATGATTTGAA 33756

```
QY 661 ACCGTTAGGATATTTTCTCACCAGCAAAATCTTGTTCAGAGTGAACACATTCCTTTT 720
Db 33757 ACTGTAGAGGTTTCTTCTTCAGATAGTCAAAATGTTATTTAGAGTGAATAATCAGCTTC 33816
QY 721 TATACAGCCTCTTAGAGGAAATTTATCCCATACAGACCGTTTATTATTAACAGAGATTT 780
Db 33817 TATACAGCTTGTCTTAGAGGAAATCTACCTCATACTGATCGCTCTTAATCAAGTTT 33876
QY 781 GAGACGGAGGTTGTTTTCATACCAATCCCTTCGCCACGCTATGGAACGTCCTTCTTG 840
Db 33877 GAAACCGAAATATCTTTAATACAAATGCTTTAGCCATGCTATGGAACGCTTATTTA 33936
QY 841 ATTTCTAATGCTACTCAAAAGGCTACTGTTAAAGCTTGAGATTCTCAAAATCATATTTCA 900
Db 33937 ATTTGGAATGCAACTCAGAACGGTACTGTTGTTTGAAGAAATTCAAAATGAGACAGTCTCA 33996
QY 901 GCTCATGTAACTCACTGAGGTTGTTAAAGGTAAGGTAAGGAGGATTTAGATATTGTTAGTCAG 960
Db 33997 GCTCATGTAACTCTCCAGAGTTGGTAAAGTTAATGAGGAATTTGGATCTGTTAGTCTT 34056
QY 961 TCTGCTAGTATTTAACTATCAGCTTCAATCCAACTTACCTTATGAGCTTTTAAAGCT 1020
Db 34057 AAAGGTGATAGTTTAAATATTAGCTTTAATCCAACTTACCTAATGAAATCTTTAAAGCA 34116
QY 1021 ATTTAAAGTGAACAGTAAATTCATTTCTTATCACCAGTTCGACCATTCACCCTAACA 1080
Db 34117 GTAAAGCAAGAACAGTATGAGTTTATTTTCTCCAGTACGTCCTTACTTTTGACA 34176
QY 1081 CCAGCGATGAGGAAGAAAGTTTATCAATTAATTAACACAGTACGAAACAAAC 1134
Db 34177 CTGCTGAAGATCTGGAAGATTTCAATTAATTAATTAATTAATTAATTAATTAATTAAT 34230

RESULT 16
LOCUS SAG766843 143050 bp DNA linear BCT 05-NOV-2002
DEFINITION Streptococcus agalactiae NEM316 complete genome, segment 1.
ACCESSION AL766843 AL766843
VERSION AL766843.1 GI:23094426
KEYWORDS
SOURCE Streptococcus agalactiae NEM316
ORGANISM Streptococcus agalactiae NEM316
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE 1
AUTHORS Glaser,P., Rusniok,C., Chevallier,P., Buchrieser,C., Frangeul,L.,
Zouine,M., Couve,E., Lalioui,L., Msadek,T., Poyart,C.,
Trieu-Cuot,P. and Kunst,P.
TITLE Genome sequence of Streptococcus agalactiae, a pathogen causing
invasive neonatal disease
JOURNAL Mol. Microbiol. 45 (6), 1499-1513 (2002)
MEDLINE 22242508
PUBMED 12354221
REFERENCE 2
AUTHORS Glaser,P., Rusniok,C. and Frangeul,L.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-2002) Glaser P., Institut Pasteur, Genomique des
Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris
Cedex 15, FRANCE. E-mail: pglaser@pasteur.fr Phone: +33 (0)1 45 68
89 96, Fax: +33 (0)1 45 68 87 86
FEATURES
source
location/Qualifiers
1..143050
/organism="Streptococcus agalactiae NEM316"
/mol_type="genomic DNA"
/strain="NEM316"
/serovar="serovar III"
/db_xref="taxon:211110"
165..1537
/gene="dnaA"
165..170
/gene="dnaA"
176..1537
/gene="dnaA"

gene
RBS
CDS
```



```

gene      4188.  .4391
          /gene="gbs0005"
CDS       4188.  .4391

```

```

/note="Hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="CAD45650.1"
/db_xref="GI:23094431"
/db_xref="TrEMBL:Q8E7Z0"
/translation="MDNLRLRAEIFDVS

```

gene	4537..5666	/gene="gbs0006"
RBS	4537..4542	/gene="gbs0006"
CDS	4551..5666	

/note="similar to hypothetical GTP binding protein"/

```

/codon_start=1
/transl_table=11
/product="Unknown"
/protein_id="CAD"
/db_xref="GI:2301"
/db_xref="GOA:Q8"
/db_xref="TREMBL

```

[illegible]

```

gene
LIQIGSEKAVKEAGRLR
join(5739..5744,
/gene="gbs0007"
5739..5744
/gene="gbs0007"
5750..6325
RBS
CDS

```

/gene="gbs0007"  
/note="similar to peptidyl-tRNA hydrolase"

```
codon_start=1
/transl_table=11
product="Unknown protein"
protein_id="CAD444"
/db_xref="GI:230999999"
/db_xref="GOA:Q8E8E8"
/db_xref="TrEMBL:Q8E8E8"
```

[illegible]

```

CDS
        /gene="gbs0008"
        6322..9819
        /gene="gbs0008"
        /note="similar to
        /codon_start=1

```

```
transl_table=11
/product="Unknown"
```

```

protein_id="CAD4
db_xref="GI:2309
db_xref="GOA:Q8E
db_xref="TREMBL:

```

tränellation="MNI:ELFSQNVVRVHWSGLVLTNRSLVWVFGSGASKAIAIASY  
TSDSDSLSSILIGDENYVFPADDAVETAFSSJDSISRL  
KLSKLIWVTPKNGVLTISGLRLLPENFVSQVKEFGEYDLKCNLNVG  
QKVSQVFPESGSGOIIDLEIMTQVYRLEFDEJIDQIDPDKSLKOLE  
QVSPADDILOADPERAKKLEGLYSLASVORTLVSTENFPHKSIDRR  
SIPYEKQGLQIDIPGTELPVDDFOKIDVRNAKLDLEATLLEDOCKSHSSIN  
FSDPYKQGLQVOPATFSPNFHNGLGNLEKHLHFQVGMQBFQVFLVDENRY  
GSGATVLLVDQSGKLANLOENKEVGLDIIISDKNDVOKESGLVHLHSGVFA  
EKTIVLTREIHRVRKRLRANSRNLKNDYSGVDYVHNVHVGKGLFE  
LEIQGHRDYLTYQVADRISIPVEQIELLTKYASGKEPKINDGDFKKAQ

RVAKQVEDIADLLKLYAERSOLOQFAGSPDDNMQNDPNDFAIVETEDQLRSIKGIB  
QOMGNRPMDLLVSDGVGKTEVAMRAKQVNDHKKVWVLVPTTLVAQOHFENFKE  
RF SNYPVTVDLSRSGKEQDTHLKLRSKGQVDIIIGHRLSLODVFSLGLIIVID  
EEORGVGHKEKELKTKVDVLTLTATPIPTLHWSMLGIRDLSVIEPTTPNRYPVQ  
TVYLTENGLVREALIREIDRGQVYVYVNRKVDITDQKVESELQELVPEASIGFVGHQM  
SEIQUENTLIDFINGDYDVLVATTIETGVDSINVTLPVENADHMLGSLTVLQGRV  
GRNSRIAYALMYRPDKVLVETISEKRLDAIKTELGSQFKIAMRDLSTRAGNINIGA  
SQSGFIDSVGFMYSQLLEOALATKQGSILKIQKNAEALOIDAYLPAEYISDERQK  
LEIYKIRBELSTRADYALQDELINDRFGEYPDQVAYLLIEGLLGLAYDLDAETLVERK  
GNEISILPEKASLKFYLDQYFEALSKTOLKFAIBSETNGKMEWFENIKHKQNYEITEE  
LLKFABCFETIKSKRPVEE"  
9893. .9909  
/gene="gbs0008"  
join(10095. .10100,10110. .10382)  
/gene="gbs0009"  
10095. .10100

Query Match	56.5%	Score 641.2	DB 1	Length 143050
Best Local Similarity	72.8%	Pred. No. 1.9e-108		
Matches 826	Conservative	0	Mismatches 308	Indels 0
	Gaps	0		

Qy	1	ATGATTCAA	TTTTCAATTAATCGC	ACATTA	TTTATTTATCGCTTTAAATCAAC	CTA	AAACGT	60
Db	1692	ATGATTCA	TTTTTCAATAAATAAAATTTCTTTC	GCATGC	TTAAGT	TAACCGGTA	ACCAACGA	1751
Qy	61	GCATTAGCA	CTAAAAATG	CCATCTCTATTTCTTTTC	ATCAATAAAAAT	TGAAGTCA	CTTCT	120
Db	1752	GCATTAGT	CATAAAAATG	CGATTCCAATCTCTTCAACTG	TTAAATAG	AAGTGA	CTAGG	1811
Qy	121	ACAGGATGA	CTTTTAACAGGGTCT	TAACGGTCAAAATATCAATTTG	AAAAACATTA	TTCTCTGT	A	180
Db	1812	GATGCTATCA	TTTTTAACGGGTC	CAATGACAAATTTCAATTTG	AAAAATCA	TTTCTCTGCT	1871	
Qy	181	AGTAATGA	AAATGCTGGTTGCTPAATTA	CTCTCCAGGAGCTATTTTA	TAGAAGCTAGT	240		
Db	1872	TCAAATGA	AAATCGAGTTTACTAGTAACGAAT	CCAGGCTCTATTTGT	TAGAAGCTG	1931		
Qy	241	TTTTTTATTA	ATAATTAATTTCAAGTTTG	CCAGATATTTAGTATAAATG	TTAAAGAAATGA	300		
Db	1932	TTCTTTATTA	ATAATTAATTTCAAGTTTAC	CAGATGAACCTTTAGAAATTTAC	CAGAGATTGA	1991		
Qy	301	CAACACCA	AGTTGTTTTTAACCACTGGT	TAAATCAGAGATTA	CTTAAAAAGGAAAAAGATGTT	360		
Db	1992	CAACATCA	AAATGTTCTTACTAGTGGAAAT	TCAGAGATTA	CTTTGAAAGGTAG	AGATGTC	2051	
Qy	361	GACCAATCT	CTCGTCTACAAGAGTA	TCACAGAAAACTCTTTTGATTTT	AAAAACAAA	420		
Db	2052	GATCAATAC	CCCTCGTGTGAGGAAATG	CAACAGATCTCCATTAACAT	TAGAACTAAA	2111		
Qy	421	TTATTCAAGT	CTATTATTGCTCAAAACAGTTT	TGCAGCCAGTTTAC	AGAAAGTCGTCCT	480		
Db	2112	CTGTTAAAT	CAATTAATTAATGAAACTG	CTTTTGTCTAGCC	ACACAGAAAGCGGTC	2171		
Qy	481	ATTTTAAC	AGGAGTTTCAATTTGTA	TTAAGTAATCATAAAGATTTT	TAAAGCAGTAGCGACT	540		
Db	2172	ATCTTAAC	AGGTGTTCAITTTGGTTATCAGT	CAAAATAATACTTTAGGCTG	TTGCCACA	2231		
Qy	541	GACTCTCA	TCGTATGAGCCAGTTTAA	TCTCTTTGGACAATA	CTTCAGCAGATTTGATG	600		
Db	2232	GATTCAC	CCGATGAGTCAACGCACTTTCCAA	TTAGAGAAATCGGCTAATA	TTTTTGAT	2291		
Qy	601	GTAGTCTT	CCAAATTAATCTTTGAGAGAA	TTTTTTCAGCAGTATTTT	TACAGATGATTTGAG	660		
Db	2292	TTGGTTGT	CCAGTAAATCCCTTCGAAATTT	TCGGCTGTTTTTACAGATG	ATTTGAA	2351		
Qy	661	ACGTTGAG	GTATTTTTTCTCAACAGCCAA	ATCTTGTTCAGAGTGA	CAATTTCTTTT	720		
Db	2352	ACTGTAGA	GAGTTTCTTCTCAGATAGT	CAAAATGTTATTTAGAGT	GAAATATCAGCTTC	2411		
Qy	721	TATACAG	CCCTCTTAGAGGAAATTA	TCCGATACAGACCGT	TTTATTAATGACAGAA	780		
Db	2412	TATACAG	GTTCGTTGAAGGAAATCA	CCGTGATCTGATCGCC	TCTTAATCACTGTT	2471		



```

QY 781 GAGACGAGGTTGTTTCAATACCAATCCCTTCGCCACGCTATGGAACGTCCTTCTTG 840
Db 2472 GAAACCGGAAATTATCTTTAATACAAATGCTTTACGCCATGCTATGGAACGTCCTTATTTA 2531
QY 841 ATTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTTGAGATTTACTCAAAAATCATATTTCA 900
Db 2532 ATTTGCAATGCACTCAGAACGGTACTGTTGTTTAGAAATTCAAAATGAGACAGTCTCA 2591
QY 901 GCTCATGTTAACTCACTGAGGTTGGTAAAGGTAACGAGGATTTAGATATTTAGTTAGTCAG 960
Db 2592 GCTCATGTTAACTCTCCAGAAAGTTGGTAAAGTAAATGAGGAATTCGATCTGTTAGTCTT 2651
QY 961 TCTGTAGTGAATTTAACTATCAGCTTCAATCCAACTTACCTTATTTAGCTCTTTTAAAGCT 1020
Db 2652 AAAGTGATAGTTTAAATATTTAGCTTTTAAATCCAACTTACCTTAAATTTGAATTTTAAAGCA 2711
QY 1021 ATTTAAAGTGAACAGTAAATTAATTTCTTTATCACCAGTTTCGACCAATTCACCCCTAACA 1080
Db 2712 GTAAAACGGAACAGTTACGATTCGATTTATTTCTCCAGTACGTCCTTACCTTTGACA 2771
QY 1081 CCAGCGATGAGGAAGAAAGTTTATCCAAATTAATACAGGATGAGCAACAAAC 1134
Db 2772 CCTGCTGAAGATACTGAAGATTTTCATCCAAATTAATAACTCTGTTGCTACTAAC 2825

RESULT 17
CQ655073 349980 bp DNA linear PAT 02-FEB-2004
LOCUS Sequence 12030 from Patent WO0234771.
DEFINITION CQ655073
ACCESSION CQ655073
VERSION CQ655073.1 GI:41687950
KEYWORDS Streptococcus agalactiae
SOURCE Streptococcus agalactiae
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.

REFERENCE
1 Telford, J., Masignani, V., Margarit y Ros, I., Grandi, G., Fraser, C. and Tettelin, H.
TITLE Nucleic acids and proteins from streptococcus groups a & b
JOURNAL Patent: WO 0234771-A 12030 02-MAY-2002;
Chiron S.p.A. (IT): THE INSTITUTE FOR GENOMIC RESEARCH (US)

FEATURES
source
location/Qualifiers
1..349980
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"
/note="seq 10967 to big: splitted in-seq 10.967, from 0.000.001 to 0.349.980-seq 12.025, from 0.300.001 to 0.649.980-seq 12.026, from 0.600.001 to 0.949.980-seq 12.027, from 0.900.001 to 1.249.980-seq 12.028, from 1.200.001 to 1.549.980-seq 12.029, from 1.500.001 to 1.849.980-seq 12.030, from 1.800.001 to 2.149.980-seq 12.031, from 2.100.001 to 2.160.266"

ORIGIN
Query Match 56.5%; Score 641.2; DB 6; Length 349980;
Best Local Similarity 72.8%; Pred. No. 1.6e-108;
Matches 826; Conservative 0; Mismatches 308; Indels 0; Gaps 0;

QY 1 ATGATTCAATTTTCAATTAATCGACATTTATTTATTCATGCTTTTAAATCAACTAAACGT 60
Db 333097 ATGATTCAATTTTCAATTAATTAATTTTCTTTCGATGCTTAACCGGTAAACCAACGA 333156
QY 61 GCTATTAGCACTAAAATGCCATTCCTATTCTTTCAATCAATAAAAATTTGAAGTCACTTCT 120
Db 333157 GCTATTAGTCATAAAAATGCGATTTCCATCTTTCAACTGTTTAAATAGAGTGAAGTACTAGG 333216
QY 121 ACAGAGGTAACCTTTAAACAGGGTCTAACCGTCAAAATATCAATTTGAAAAACACTATTCTCTGA 180
Db 333217 GATGCTATCAATTTAAACGGGGTCAAAATGGAACAAATTTCAATTTGAAAAATCTATTCTCTGCT 333276
QY 181 AGTAATGAAATGCTGGTTTCTGTAATTTACCTCTCAGGAGCTATTTTATTAGAGCTAGT 240

```

```

Db 333277 TCAAAATGAAATGAGGTTTACTAGTAACGAATCAAGCTCTATTTTGTTAGAGCTGCT 333336
QY 241 TTTTATTAATATATTTTCAAGTTTGGCAGATATTTAGTATATAAATGTTTAAAGAAATGAA 300
Db 333337 TTCTTTATTAATATATTTTCAAGTTTACCAGATGTAACCTTTAGAAATTTACAGAGATTGAA 333396
QY 301 CAACACCAAGTTGTTTAAACCAGTGGTAAATCAGAGATTACCTTTAAAGGAAAGATGTT 360
Db 333397 CAACATCAAAATGTTTCTTACTAGTGGAAATCAGAGATTACTTTTGAAGAAAGATGTC 333456
QY 361 GACCAGTATCTCTGTACAAAGAAAGTATCAACAGAAAAATCCTTTGATTTTAAAAACAAA 420
Db 333457 GATCAATACCTCTGTTCGAGAAATGACAACAGATCTCCATTAACATTTAGAAACTAAA 333516
QY 421 TTATTGAAGTCTATTTATTTGCTGAAACAGCTTTTGCAGCCAGTTTCAAGAAAGTCTGCT 480
Db 333517 CTGTTAAAAATCAATTTATTAATGAAATGCTTTTGTCTAGCAACAAGAAAGCGTCCA 333576
QY 481 ATTTTAAACAGGAGTTCATATTTGATTAAGTAATCATTAAGATTTTAAAGCAGTAGCGACT 540
Db 333577 ATCTTAAACAGGTGTTCAATTTGCTTTATCAGTCAAAAATAAATACTTTAAGGCTGTTGCGACA 333636
QY 541 GACTCTCATCTGATGAGCCAACTTTAATCACTTTGGACAATCTTCAAGCAGATTGTATG 600
Db 333637 GATTCAACCCGTATGAGTCAACGCACTTTCCAAATTAGAGAAATCGGCTAATAATTTTGAAT 333696
QY 601 GTAGTTCTTCCAAGTAAATCTTTGAGAGAAATTTTTCAGCAGTATTTTACAGATGATATTGAG 660
Db 333697 TTGGTTGTTCCAAGTAAATCTCTTCGAGAAATTTTTCGGCTGTTTTTACAGATGATATTGAA 333756
QY 661 ACCGTTGAGGTATTTTCTCACCAGCCAAATCTTTGTCAGAGTGAACACATTTCTTTT 720
Db 333757 ACTGTAGAGGTTTCTTCTCAGATAGTCAAAATGTTTATTAGAGTGAAAAATATCAGCTTC 333816
QY 721 TATACAGCCTCTTAGAAGGAAATTTATCCCGATACAGACCGTTTATTATATGACAGAAATTT 780
Db 333817 TATACAGCTTTGTTGAAGGAAATCTACCTCGATCTGATCGCTCCTTAATCACTGTTT 333876
QY 781 GAGACGAGGTTGTTTTCATACCCAAATCCCTTCGCCACGCTATGGAACGTCCTTCTTTG 840
Db 333877 GAAACCGAAATTAATCTTTTAAATACAAATGCTTTTACGCCATGCTATGGAACGCTTATTTA 333936
QY 841 ATTTCTAATGCTACTCAAAATGCTACTGTTAAGCTTTGAGATTACTCAAAAATCATATTTCA 900
Db 333937 ATTTGCAATGCAACTCAGAACGCTACTGTTCTGTTTAGAAATTTCAAAAATGAGACAGTCTCA 333996
QY 901 GCTCATGTTAACTCACTGAGGTTGGTAAAGTAAAGTAAATGAGGAATTCGATCTGTTAGTCTT 960
Db 333997 GCTCATGTTAACTCTCCAGAAAGTTGGTAAAGTAAATGAGGAATTCGATCTGTTAGTCTT 334056
QY 961 TCTGTAGTGAATTTAACTATCAGCTTCAATCCAACTTACCTTATTTAGTCTTTTAAAGCT 1020
Db 334057 AAAGTGATAGTTTAAATATTTAGCTTTAATCCAACTTACCTTAAATTTGAATCTTTAAAGCA 334116
QY 1021 ATTTAAAGTGAACAGTAAATTTCAATTTCTTATCACCAGTTCGACCAATTCACCTTACA 1080
Db 334117 GTAAAAACGGAACAGTTACGATTTCAATTTTATTTCTCCAGTACGTCCTTACTTTTGACA 334176
QY 1081 CCAGCGATGAGGAAGAAAGTTTATCCAAATTAATTAACACAGTACGCAACAAAC 1134
Db 334177 CCGTGTGAAGATCTGAAGATTTCAATTTCAATTAATTAATCTCTGTTGCTACTAAC 334230

```

```

RESULT 18
AX954527 349980 bp DNA linear PAT 08-JAN-2004
LOCUS Sequence 1373 from Patent WO03093306.
DEFINITION AX954527
ACCESSION AX954527
VERSION AX954527.1 GI:40783900
KEYWORDS Streptococcus pyogenes
SOURCE Streptococcus pyogenes
ORGANISM Streptococcus pyogenes

```

```
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1
Telford, J., Massignani, V., margarit y Ros, I., Grandi, G., Fraser, C.
and Tettelin, H.
Nucleic acids and proteins from streptococcus groups a b
Patent: WO 03093306-A 1373 13-NOV-2003;
Chiron SpA (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
Location/Qualifiers
1. 349980
/organism="Streptococcus pyogenes"
/mol_type="unassigned DNA"
/db_xref="taxon:1314"
/notes="seq 1373 to long, original length:
2.160.266-replaced by-seq 1373: from 0.000.001 to
0.349.980-seq 1374: from 0.300.001 to 0.649.980-seq 1375:
from 0.600.001 to 0.949.980-seq 1376: from 0.900.001 to
1.249.980-seq 1377: from 1.200.001 to 1.549.980-seq 1378:
from 1.500.001 to 1.849.980-seq 1379: from 1.800.001 to
2.149.980-seq 1380: from 2.100.001 to 2.160.266"
ORIGIN
Query Match 56.5%; Score 641.2; DB 6; Length 349980;
Best Local Similarity 72.8%; Pred. No. 1.6e-108;
Matches 826; Conservative 0; Mismatches 308; Indels 0; Gaps 0;
QY 1 ATGATTCAATTTTCAATTAATCGACATTTATTTATTCATGCTTTAAATCAACTAAACGT 60
Db 1618 ATGATTCAATTTTCAATTAATTAATAAATTTCTTCTGATGCTTAACGTAACCAACGA 1677
QY 61 GCTATTAGCACTAAATGCAATTCCTATTCTTCATCAATTAATAATGAAGTCACTTCT 120
Db 1678 GCTATTAGTCAATAAATGGATTCCTCAATCCCTTTCAACTGTTAAATAAGAGTACTAGG 1737
QY 121 ACAGGAGTAACTTTAAACAGGCTTAACGGTCAAAATCAATTAATGAAACACATTTCTGTA 180
Db 1738 GATGCTATCATTTTAAACGGGTCAAATGACAAATTCATTTGAATACTATTCTGCT 1797
QY 181 AGTAATGAAATGCTGGTTGCTAATTAATCTCTCCAGAGCTATTTTATTAAGAGCTAGT 240
Db 1798 TCMAATGAAATGCGAGTTTACTAGTAACGAATCCAGGCTCTATTCTTGTAGAGCTGGT 1857
QY 241 TTTTTPATTAATTAATTTCAAGTTTGCAGATATTAGTATTAATGTTAAAGAAATTTGAA 300
Db 1858 TTCTTTATTAATTAATTTTCAAGTTTACCAGATGTAATTTAGAAATTTACAGAGATTTGAA 1917
QY 301 CAACACCAATTTGTTTAAACAGTGGTAAATCAGAGATTACCTTAAAGGAAAGAGATGTT 360
Db 1918 CAACATCAATTTGTTTCTTACTAGTGGAAATCAGAGATTACTTTGAAAGGTAAAGATGTC 1977
QY 361 GACCATATCTCTGCTCAAGAGATATCAACAGAAATCTTTGATTTTAAACAAATA 420
Db 1978 GATCAATACCTCTGCTGAGGAATGACACAGATCTCCATTAACATTAAGAACTAAA 2037
QY 421 TTATTGAAGTCTATTATTGCTGAAACAGCTTTTGCAGCAGTTTACAGAAAGTCGCTCT 480
Db 2038 CTGTATAAATCAATTAATTAATGAAATGCTTTGCTGCTAGCCCAACAGAAAGCGTCCA 2097
QY 481 ATTTTAAACAGAGTTTCAATTTAGTAAATCATTAAGATTTTAAAGCAGTAGCAGT 540
Db 2098 ATCTTAAACAGGTTTCAATTTAGTAAATCATTAAGATTTTAAAGCAGTAGCAGT 540
QY 541 GACTCTCATCTGATGAGCCAAAGCTTTAATCACTTTGGCAATATCTTCCAGCAGATTTGATG 600
Db 2158 GATTACACACCTGATGAGTCAACGCACTTTCCAAATTAGAGAAATCGGTAATATTTTGTAT 2217
QY 601 GTAGTTCTTCAAGTAAATCTTTGAGAGAAATTTTACAGATTTTACAGATGATTTAG 660
Db 2218 TTGGTTGTTCCAGTAAATCTCTTCGAGAAATTTTCGGCTCTTTTACAGATGATTTGAA 2277
QY 661 ACCGTTGAGTATTTTCTCACCAGCCAAATCTTGTTCAGAGAGTGAACATTTCTTTT 720
Db 2278 ACTGAGAGGTTTCTCTCTCAGATAGTCAATGTTTATTAGAGTGAATATCAGCTTC 2337
QY 721 TATACAGCCTCTTAGAAGGAATTTATCCGATACAGACCGTTTATTATGACAGAAATTT 780
Db 2338 TATACAGCTTTGCTTTGAAGAAATCTCCCTGATGATCGCTCTTAATCAATCAGTTT 2397
QY 781 GAGACGGAGTTGTTTCAATACCAATCCCTTCGCCACGCTATGGAACGCTCTTCTTGTG 840
Db 2398 GAAACCGAAATTTATCTTTAATACAAATGCTTTACGCCATGCTATGGAACGCTTATTATA 2457
QY 841 ATTTCTAATGCTACTCAAAATGCTTAAAGCTTTAGATTTGAGATTTCAAAATCATATTTCA 900
Db 2458 ATTTGGAATGCAATCAGAACGTAATCTGTTGTTTGAATAATCAAAATGACAGAGTCTCA 2517
QY 901 GCTCATGTTAACTCACTGAGGTTGTAAGTAAGCAAGGATTTAGATTTGTTAGTCTAG 960
Db 2518 GCTCATGTTAACTCTCCAGAGTTGTAAGTAAGTAAATGAGGAATTTAGTCTTGTAGTCTT 2577
QY 961 TCTGCTAGTAACTTAACTATCAGCTTCAATCAACTTACCTTATTTAGAGTCTTTAAAAAGCT 1020
Db 2578 AAAGTGATAGTTTAAATATTTAGCTTTAATCCAACTTACCTAATTTGAATCTTTAAAAAGCA 2637
QY 1021 ATTAAAGTGAACAGATAAATAATTCATTTCTTATCACCAGTTGCAACATTTCACTCAACA 1080
Db 2638 GTAAAGCGAAACAGTTACGATTCGATTTATTTCTCCAGTACGTCATTTACTTTTGACA 2697
QY 1081 CCAGCGATGAGGAAGAAAGTTTATCCAAATTAATTAACAGCAGTACGAAACAAAC 1134
Db 2698 CCTGTTGAAGATACTGAAGATTTTATCCAAATTAATAACTCTCTGTTCTGACTTAC 2751
RESULT 19
AX571491 1134 bp DNA linear PAT 29-NOV-2002
LOCUS
DEFINITION Sequence 4703 from Patent WO02077021.
ACCESSION AX571491
VERSION AX571491.1 GI:26003814
KEYWORDS
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1
Massignani, V., Tettelin, H. and Fraser, C.
Streptococcus pneumoniae proteins and nucleic acids
Patent: WO 02077021-A 4703 03-OCT-2002;
Chiron SpA (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
Location/Qualifiers
1. .1134
/organism="Streptococcus pneumoniae"
/mol_type="unassigned DNA"
/db_xref="taxon:1313"
ORIGIN
Query Match 53.6%; Score 608.2; DB 6; Length 1134;
Best Local Similarity 71.1%; Pred. No. 5.9e-102;
Matches 805; Conservative 0; Mismatches 328; Indels 0; Gaps 0;
QY 1 ATGATTCAATTTTCAATTAATCGCAGATTTATTTATTCATGCTTTTAAATCAACTTAAACGT 60
Db 1 ATGATTCAATTTTCAATTAATTAATAAATTTATTTCTCAAGCATTAATAACTACTACTAAGAGA 60
QY 61 GCTATTAGCATAAATAATGCCATTCCTATTCTTTTCATCAATAAATAATTTGAAGTCACTTCT 120
Db 61 GCTATTAGTCTTAAATAATGCCATTCCTATTCTTTATCAACAGATAAATAATTTGACGTGACCAAT 120
QY 121 ACAGAGTAACTTTAAACAGGCTTAAACGTCAAATATCAATTAATGAAACACTATTTCTGTA 180
Db 121 GAAGTATTTACTTTTAAATTTGGTTTCAATTTCAATTTGAAATTTTATTTTCTCTCAA 180
QY 181 AGTAATGAAATGCTGGTTTCTGCTAATTTACCTCTCTCAGAGCTATTTTATTAGAGAGTAGT 240
Db 181 AAAAATGAAGATGCTGGTTTGTAAATTTCTTTTGTAGTCTCTTCTTCTGAGTCTCTTCTTGAAGCTTCT 240
```

```
QY 241 TTTTATTATTAATATTATTTTCTAGTTTGCAGATATTAGTATAAATGTTTAAAGAAATTCGA 300
Db 241 TTTTATTATTAATATTATTTCTAGTTTGCAGATATTAGTATAAATGTTTAAAGAAATTCGA 300
QY 301 CAACACCAAGTTGTTTAAACCAAGTGGTAAATCAGAGATTACCTTAAAGGAAAGATGTT 360
Db 301 CAATATCAATTTGTTTAAACCAAGTGGTAAATCAGAGATTACCTTAAAGGAAAGATGTT 360
QY 361 GACAGTATCTCGTCTCAAGAGATATCAACAGAAATTCCTTTGATTTTAAACCAAAA 420
Db 361 GAACAATATCCAGCAATCAAGAAATTTTCAAGCAAGTCTCTTTTAAATCTTTGAAACAAA 420
QY 421 TTATTGAAGTCTATTATTGCTGAACAGCTTTTGCAGCCAGTTTCAAGAAAGTCTGCT 480
Db 421 TTACTCAAGAAATTTATTATGAACAGCTTTTGCAGCCAGTTTCAAGAAAGTCTGCT 480
QY 481 ATTTTAAACAGGAGTTTCAATTTGATTTAAGTATATCAATAAGATTTTAAAGCAGTACGACT 540
Db 481 ATTTTAAACAGGAGTTTCAATTTGATTTAAGTATATCAATAAGATTTTAAAGCAGTACGACT 540
QY 541 GACTCTCATCGTATGAGCCAGTTTAAATCACTTTTGGCAATATCTTCAAGCAGATTGATG 600
Db 541 GACTCTCATCGTATGAGCCAGTTTAAATCACTTTTGGCAATATCTTCAAGCAGATTGATG 600
QY 601 GTAGTTCTTCCAGTAAATCTTTGAGAGATTTTCCAGCAGTATTTACAGATCATATTGAG 660
Db 601 GTAGTTCTTCCAGTAAATCTTTGAGAGATTTTCCAGCAGTATTTACAGATCATATTGAG 660
QY 661 ACCGTGTAGGATTTTCTTCCAGGCAAAATCTTTGTTGAGAGTGAACACATTTCTTTT 720
Db 661 ACTGTAGAGATTTTCTTCCAGGCAAAATCTTTGTTGAGAGTGAACACATTTCTTTT 720
QY 721 TATACAGCGCTTTAGAGGAAATTTATCCGATACAGACCGTTTATTAATGACAGAAATTT 780
Db 721 TATACAGCGCTTTAGAGGAAATTTATCCGATACAGACCGTTTATTAATGACAGAAATTT 780
QY 781 GAGCGGAGTTGTTTCAATACCAATCCCTTCCGCAAGTATCGAGAGTGGCTTCTT 840
Db 781 AACACTACTATTACTTTTAAATGTTGTTTCAACCAAGTATCGAGAGTGGCTTCTT 840
QY 841 ATTTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTGGAGTTTACCTCAAAATCATATTTCA 900
Db 841 TTATCAAGTGGACTCAAAATGGTACTGTTAAGCTTGGAGTTTACCTCAAAATCATATTTCA 900
QY 901 GCTCATGTTAATCACTGAGGTTGGTAAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGG 960
Db 901 GCTCATGTTAATCACTGAGGTTGGTAAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGG 960
QY 961 TCTGCTAGTATTTAATCACTGAGTTCATCCAGTTCCTTATGAGTCTTTAAAGCT 1020
Db 961 ACTGCTAGGATTTTGAACATTTAGTTTCAACCAAGTATCGAGAGTGGCTTCTT 1020
QY 1021 ATTAAGTGAACAGTAAATTTCAATTTCTTATCAGCAGTTTCCAGCAGTTCACCCCTAACA 1080
Db 1021 TTAATAGGAAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGG 1080
QY 1081 CCAGCGATAGGAAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGG 1133
Db 1081 CCAGCGATAGGAAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGG 1133
```

## RESULT 20

```
AR480840
LOCUS 1137 bp DNA linear PAT 14-MAY-2004
DEFINITION Sequence 1744 from patent US 6699703.
ACCESSION AR480840
VERSION AR480840.1 GI:47239802
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1137)
AUTHORS Doucette-Stamm,L., Bush,D., Zeng,Q., Opperman,T. and
```

```
Houseweart, C.E.
Nucleic acid and amino acid sequences relating to Streptococcus
pneumoniae for diagnostics and therapeutics
Patent: US 6699703-A 1744 02-MAR-2004;
Location/Qualifiers
1. 1137
/organism="unknown"
/mol_type="genomic DNA"

Query Match 53.6%; Score 608.2; DB 6; Length 1137;
Best Local Similarity 71.1%; Pred. No. 5.9e-102;
Matches 805; Conservative 0; Mismatches 328; Indels 0; Gaps 0;

ORIGIN
QY 1 ATGATTCATTTTCAATTAATCGCACATTTATTTATTCATGCTTTTAAATACAACTAAACGT 60
Db 1 ATGATTCATTTTCAATTAATTAATAATTTTATTTCTACAGCATTAATACTACTAAGAGA 60
QY 61 GCTATTAGCACTAAATGCGCATTTCTTATTTTCATCAATAAATAATGAACTCACTTCT 120
Db 61 GCTATTAGCTTAAATGCGCATTTCTTATTTTCAACAGTAAATAATGACGTGACCAAT 120
QY 121 ACAGGAGTAATTTAAACAGGCTCTAACGGTCAAAATATCAATTTGAAAAACACTATTCTGTA 180
Db 121 GAAGGTATTAATTTGTTTCAATGGTCAAAATTTCAATTTGAAAAATTTATTTCTCAA 180
QY 181 AGTAATGAAATGCTGTTTCTAATTAACCTCTCCAGGAGCTATTTTATTAAGAGCTAGT 240
Db 181 AAAAATGAAGTCTGTTTGTAAATTAATTTCTTTAGGTTGATCTCTTTGAAAGCTTCT 240
QY 241 TTTTATTATATATTTTCAAGTTTGCAGATATTAGTATAAATGTTTAAAGAAATTCGA 300
Db 241 TTTTATTCAATAGTATCTAGTTTACCTGATGAATCTTTGATTTTAAAGAAATTCGA 300
QY 301 CAACCAAGTTGTTTAAACCAAGTGGTAAATCAGAGATTACCTTAAAGGAAAGATGTT 360
Db 301 CAATCAATTTGTTTAAACCAAGTGGTAAATCAGAGATTACCTTAAAGGAAAGATGTT 360
QY 361 GACAGTATCTCGTCTCAAGAGATATCAACAGAAATTCCTTTGATTTTAAACCAAAA 420
Db 361 GAACAATATCCAGCAATCAAGAAATTTTCAAGCAAGTCTCTTTTAAATCTTTGAAACAAA 420
QY 421 TTATGAGTCTATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 480
Db 421 TTACTCAAGAAATTTTAAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 480
QY 481 ATTTTAAACAGGTTTCAATTTGATTTAAGTATATCAATAAGATTTTAAAGCAGTACGACT 540
Db 481 ATTTTAAACAGGTTTCAATTTGATTTAAGTATATCAATAAGATTTTAAAGCAGTACGACT 540
QY 541 GACTCTCATCGTATGAGCCAGTTTAAATCACTTTTGGCAATATCTTCAAGCAGATTGATG 600
Db 541 GACTCTCATCGTATGAGCCAGTTTAAATCACTTTTGGCAATATCTTCAAGCAGATTGATG 600
QY 601 GTAGTTCTTCCAGTAAATCTTTGAGAGATTTTCCAGCAGTATTTACAGATCATATTGAG 660
Db 601 GTGCTAATTTCTAGCGGTTCTTCAAGCAAGTCTTCTTCAAGCAAGTCTTCTTCAAGCAAG 660
QY 661 ACCGTGTAGGATTTTCTTCCAGGCAAAATCTTTGTTGAGAGTGAACACATTTCTTTT 720
Db 661 ACTGTAGAGATTTTCTTCCAGGCAAAATCTTTGTTGAGAGTGAACACATTTCTTTT 720
QY 721 TATACAGCGCTTTAGAGGAAATTTATCCGATACAGACCGTTTATTAATGACAGAAATTT 780
Db 721 TATACTGCTCTTCAAGGAAATTTTCAAGCAAGTATCGAGAGTGGCTTCTT 780
QY 781 GAGCGGAGTTGTTTCAATACCAATCCCTTCCGCAAGTATCGAGAGTGGCTTCTT 840
Db 781 AACACTACTATTACTTTTAAATGTTGTTTCAACCAAGTATCGAGAGTGGCTTCTT 840
QY 841 ATTTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTGGAGTTTACCTCAAAATCATATTTCA 900
Db 841 TTATCAAGTGGACTCAAAATGGTACTGTTAAGCTTGGAGTTTACCTCAAAATCATATTTCA 900
QY 901 GCTCATGTTAATCACTGAGGTTGGTAAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGG 960
Db 901 GCTCATGTTAATCACTGAGGTTGGTAAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGG 960
QY 961 TCTGCTAGTATTTAATCACTGAGTTCATCCAGTTCCTTATGAGTCTTTAAAGCT 1020
Db 961 ACTGCTAGGATTTTGAACATTTAGTTTCAACCAAGTATCGAGAGTGGCTTCTT 1020
QY 1021 ATTAAGTGAACAGTAAATTTCAATTTCTTATCAGCAGTTTCCAGCAGTTCACCCCTAACA 1080
Db 1021 TTAATAGGAAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGG 1080
QY 1081 CCAGCGATAGGAAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGG 1133
Db 1081 CCAGCGATAGGAAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGG 1133
```

Qy	901	GCTCATGTTAACTCACCTGGAGTGTGGTAAGGATAACGAGGATTTAGATATTGTTAGTCAG	960
Dd	901	GCCCATGTTCACTCTCCAGAAGTTGGTAAGTAAACGAAGAAATCGAATCTGATCAGGTT	960
Qy	961	TCTGGTAGTGATTTAACTATCAGCTTCAATCCAATCTTACTTATGTAGTCTTTAAAAGCT	1020
Dd	961	ACTGGTGAAGATTGACCATTAGTTTCAACCACCTTACTTGTGATGATTCTCTTAAAGCT	1020
Qy	1021	ATTTAAAGTGAACACGATAAAAATTCATTCTTATCACCAGTTCGACCATTCACCCYACA	1080
Dd	1021	TTAAATAGCGAAAGGTGACCATTAGCTTTATCTCAGCTGTTGTCCTTACTCTCTGTG	1080
Qy	1081	CCAGGCATGAGGAAGAAAGTTTTATCCAAATTAATACACGATGACCAAAA	1133
Dd	1081	CCAGCAGATACCTGACGAAGACTTCATGCAGCTCATTAACACAGTTGTCACAAA	1133

  

RESULT	21
AE007318	
LOCUS	
DEFINITION	Streptococcus pneumoniae TIGR4 section 1 of 194 of the complete genome.
ACCESSION	AE007318 AE005672
VERSION	AE007318.1 GI:14971464
KEYWORDS	
SOURCE	Streptococcus pneumoniae TIGR4
ORGANISM	Streptococcus pneumoniae TIGR4
REFERENCE	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
AUTHORS	1 (bases 1 to 14418) Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D., Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J., Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D., Umayam,L.A., White,O., Salzberg,S.L., Lewis,M.R., Radune,D., Holtzapple,E., Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L., McDonald,L.A., Feldblyum,T.V., Angiuoli,S., Dickinson,T., Hickey,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O., Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M. Complete genome sequence of a virulent isolate of Streptococcus pneumoniae Science 293 (5529), 498-506 (2001)
JOURNAL	Science 293 (5529), 498-506 (2001)
MEDLINE	21357209
PUBMED	11463916
REFERENCE	2 (bases 1 to 14418) Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D., Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J., Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D., Umayam,L.A., White,O., Lewis,M.R., Radune,D., Holtzapple,E., Khouri,H., Wolf,A.M., Utterback,T.R., Hansen,C.L., McDonald,L.A., Feldblyum,T.V., Angiuoli,S., Gesuwani,P., Hickey,E.K., McDonald,L.A., Loftus,B.J., Ujwal,M.L., Yang,F., Smith,H.O., Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M. Direct Submission Submitted (29-JUN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES	Location/Qualifiers
source	1..14418 /organism="Streptococcus pneumoniae TIGR4" /mol_type="genomic DNA" /strain="TIGR4" /db_xref="taxon:170187" 197..1558 /gene="SP0001" 197..1558 /gene="SP0001" /note="identified by match to TIGR protein family HMM TIGR00362"
gene	
CDS	/codon_start=1 /transl_table=11 /product="chromosomal replication initiator protein DnaA" /protein_id="AAK74194.1"



QY 1021 ATTAAGTGAACAGTAAATTCATTTCTATCACAGTTGACCAATTCACCCCTAACA 1080  
DB |||||  
QY 2737 TTAATAGCGAAAGGTGACTATTAGCTTTATCTCAGCTGTCCTCATTTCTCTGTG 2796  
DB |||||  
QY 1081 CCAGCGATGAGGAAGAAGTTTATCCAAATTAATACACAGTACGAACAAA 1133  
DB |||||  
QY 2797 CCAGCAGATACTGACGAAGACTTCATGCACTCTATTACACAGTTCGTACAAA 2849  
DB |||||  
RESULT 22  
CQ788929 21338 bp DNA linear PAT 29-MAR-2004  
LOCUS Sequence 20 from Patent EP1400592.  
ACCESSION CQ788929  
VERSION CQ788929.1 GI:45822497  
KEYWORDS unidentified  
SOURCE unclassified.  
ORGANISM  
REFERENCE  
AUTHORS Kunsch, C.A., Choi, G.H., Dillon, P.J., Rosen, C.A., Barash, S.C.,  
Fannon, M. and Dougherty, B.A.  
TITLE Streptococcus pneumoniae polynucleotides and sequences  
JOURNAL Patent: EP 1400592-A 20 24-MAR-2004;  
HUMAN GENOME SCIENCES, INC. (US)  
FEATURES  
source Location/Qualifiers  
1. 21338  
/organism="unclassified"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"  
ORIGIN  
Query Match 53.6%; Score 608.2; DB 6; Length 21338;  
Best Local Similarity 71.1%; Pred. No. 3.2e-102;  
Matches 805; Conservative 0; Mismatches 328; Indels 0; Gaps 0;  
QY 1 ATGATTCATTTTCAATTAATGCGCATTTATTTATTCATGCTTTAAATACACTAAAGCT 60  
DB ATGATTCATTTTCAATTAATTAATAATTTATTTCTACAAGCAATTAATACTACTAAGAGA 7135  
QY 61 GCTATTAGCACTAAATGCCATTCCTATTTCTTCATCAATAAATAATGAACTCACTTCT 120  
DB |||||  
QY 7136 GCTATTAGTCTAAATGCCATTCCTATTTTATCAACAGTAATAATTTGACGACCAAT 7195  
DB |||||  
QY 121 ACAGAGTAACCTTACAGGCTCTACGCTCAATATCAATTTGAAACACATTTCTCTGA 180  
DB |||||  
QY 7196 GAAGGTATTACTTTAATGGTTCAAAATGGTCAAAATTTCAATTTGAAATTTTATTCTCA 7255  
DB |||||  
QY 181 AGTAATGAAATGCTGGTTGCTAATTAATCTCTCAGAGCTATTTTATTAGAGCTAGT 240  
DB |||||  
QY 7256 AAAATGAAGATGCTGGTTGTTAATTAATCTTTAGTTGATCTCTTTGAAGCTTCT 7315  
DB |||||  
QY 241 TTTTATTATTAATTTATTTCAAGTTTGCAGATATTAGTATAAATGTTTAAAGAAATGAA 300  
DB |||||  
QY 7316 TTCTTTATCAATGTAGTACTAGTTTACCTGATGTAATCTTGTATTTTAAAGAAATGAA 7375  
DB |||||  
QY 301 CAACACCAAGTTGTTTAAACAGCTGTAATCAATCAGAGATTACTTAAAGGAAAGATGTT 360  
DB |||||  
QY 7376 CAATATCAATGTTTAAACAGCTGTAATCAATGTAATTAATCAATGTAATTAATCAATG 7435  
DB |||||  
QY 361 GACAGTATCTCTGCTCAAGAAGTATCAACAGAAATCTCTTTGATTTTAAACAAAAA 420  
DB |||||  
QY 7436 GAACAATATCCCAAGTAATCAAGAAATTTTCAAGAGCACTCTTTAATATTGTAACAAAA 7495  
DB |||||  
QY 421 TTATTGAAGTCTATTATTGCTGAACAGCTTTTGCAGCAGCTTTTCAAGAAAGTCGTCCT 480  
DB |||||  
QY 7496 TTACTCAAGAAATTAATTAATGAACAGCTTTTCTGCAAGTACCAAGAGAGTCGTCCT 7555  
DB |||||  
QY 481 ATTTTAAACAGAGTTCAATTTGATTAATAGTAATCATAAAGATTTTAAAGCAGTAGGCACT 540  
DB |||||  
QY 7556 ATTTTAAACAGGTGCCACTTCTGATTAGTCAACACAAAGATTTAAACAGTTGCAACA 7615  
DB |||||

QY 541 GACTCTCATCGTATGAGCAACGTTTAATCACTTTTGGCAATATCTTTCAGAGATTTGATG 600  
DB |||||  
QY 7616 GACTCTCATCGCTAAGCGAGAAAAATTTGACTCTTGAAGAAAAATAGTATGATTTGAT 7675  
DB |||||  
QY 601 GTAGTTCTTCCAAAGTAATCTTTTGAAGAAATTTTTCAGAGTATTTTACAGATGATTTGAG 660  
DB |||||  
QY 7676 GTCTGAATTTCCAGCGGTTCTCTACGCGAATTTTTCAGCGGTTTTCAGAGATGATTCGAA 7735  
DB |||||  
QY 661 ACCGTGAGGTATTTTCTCCAAAGCAAAATCTTTTCAAGTGAACACATCTTTT 720  
DB |||||  
QY 7736 ACTGTAGAGATTTTCTTTGCAATTAACCAATCTCTTTTGAAGCGAAAAATTTAGCTTC 7795  
DB |||||  
QY 721 TATACACGCTCTTGAAGAAATTTATCCGATACAGACCGTTTATTAATACAGAAATTT 780  
DB |||||  
QY 7796 TATACTGCTCTCTAGAAGAAATCTCTCTGATACAGATCGCTTGTATTTCCAAACAGACTTT 7855  
DB |||||  
QY 781 GAGACGAGGTGTTTCTCAATACCAATCCCTCGCACGCTATGGACGCTCTTTG 840  
DB |||||  
QY 7856 AACACTACTATTACTTTTAAATGTGTAACTTACCCAGTCAATGGAGCGTCCGCTTT 7915  
DB |||||  
QY 841 ATTCTAATGCTACTCAAAATGGTACTGTTAAGCTTTGAGATTACTCAAAATCATATTTCA 900  
DB |||||  
QY 7916 TTATCAAGTGCAGCTCAAAATGGTACTGTGAACTTGAATTTAAGGATGGGTTGTTAGC 7975  
DB |||||  
QY 901 GCTCATGTTAACTCACCTGAGGTTGTTAAGTAAACGAGATTTAGATATTTGTTAGTCAG 960  
DB |||||  
QY 7976 GCCCATGTTCACTCTCAGAAAGTTGGTAAAGTAAACGAGAAATCGATCTGATCAGGTT 8035  
DB |||||  
QY 961 TCTGTTAGTATTTAACTTATCAGCTTCAATCCAACTTACCTTATTCAGCTTCGACCACTCA 1020  
DB |||||  
QY 8036 ACTGTTGAAGATTTGACCAATTTAGTTTCAACCACTTACTTGTGATTTCTTAAAGCT 8095  
DB |||||  
QY 1021 ATTAAGTGAACAGTAAATAATTTTCTTATTCAGCTTCGACCACTTCACCTTAACA 1080  
DB |||||  
QY 8096 TTAATAGCGAAAGGTGACTATTAGCTTTATCTCAGCTGTTCTGCTCATTTACTCTGTG 8155  
DB |||||  
QY 1081 CCAGCGATGAGGAAGAAGTTTATCCAAATTAATACACAGCTCATTCACCACTTCTGACAAA 1133  
DB |||||  
QY 8156 CCAGCAGATACTGACGAAGACTTCAATGCTGAGCTCATTCACCACTTCTGACAAA 8208  
DB |||||  
RESULT 23  
AR218788 21338 bp DNA linear PAT 25-SEP-2002  
LOCUS Sequence 20 from patent US 6420135.  
DEFINITION AR218788  
ACCESSION AR218788  
VERSION AR218788.1 GI:23319722  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 21338)  
AUTHORS Kunsch, C.A., Choi, G.H., Dillon, P.S., Rosen, C.A., Barash, S.C.,  
Fannon, M.R. and Dougherty, B.A.  
TITLE Streptococcus pneumoniae polynucleotides and sequences  
JOURNAL Patent: US 6420135-A 20 16-JUL-2002;  
FEATURES  
source Location/Qualifiers  
1. 21338  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 53.6%; Score 608.2; DB 6; Length 21338;  
Best Local Similarity 71.1%; Pred. No. 3.2e-102;  
Matches 805; Conservative 0; Mismatches 328; Indels 0; Gaps 0;  
QY 1 ATGATTCATTTTCAATTAATGCGCATTTATTTATTCATGCTTTAAATACACTAAAGCT 60  
DB ATGATTCATTTTCAATTAATTAATAATTTATTTCTACAAGCAATTAATACTACTAAGAGA 7135  
QY 61 GCTATTAGCACTAAATGCCATTCCTATTTCTTCATCAATAAATAATGAACTCACTTCT 120  
DB ATGATTCATTTTCAATTAATTAATAATTTATTTCTACAAGCAATTAATACTACTAAGAGA 7195  
QY 7136 GCTATTAGTCTAAATGCCATTCCTATTTTATCAACAGTAATAATTTGACGACCAAT 7195  
DB |||||  
QY 121 ACAGAGTAACCTTACAGGCTCTACGCTCAATATCAATTTGAAACACATTTCTCTGA 180  
DB |||||  
QY 7196 GAAGGTATTACTTTAATGGTTCAAAATGGTCAAAATTTCAATTTGAAATTTTATTCTCA 7255  
DB |||||  
QY 181 AGTAATGAAATGCTGGTTGCTAATTAATCTCTCAGAGCTATTTTATTAGAGCTAGT 240  
DB |||||  
QY 7256 AAAATGAAGATGCTGGTTGTTAATTAATCTTTAGTTGATCTCTTTGAAGCTTCT 7315  
DB |||||  
QY 241 TTTTATTATTAATTTATTTCAAGTTTGCAGATATTAGTATAAATGTTTAAAGAAATGAA 300  
DB |||||  
QY 7316 TTCTTTATCAATGTAGTACTAGTTTACCTGATGTAATCTTGTATTTTAAAGAAATGAA 7375  
DB |||||  
QY 301 CAACACCAAGTTGTTTAAACAGCTGTAATCAATCAGAGATTACTTAAAGGAAAGATGTT 360  
DB |||||  
QY 7376 CAATATCAATGTTTAAACAGCTGTAATCAATGTAATTAATCAATGTAATTAATCAATG 7435  
DB |||||  
QY 361 GACAGTATCTCTGCTCAAGAAGTATCAACAGAAATCTCTTTGATTTTAAACAAAAA 420  
DB |||||  
QY 7436 GAACAATATCCCAAGTAATCAAGAAATTTTCAAGAGCACTCTTTAATATTGTAACAAAA 7495  
DB |||||  
QY 421 TTATTGAAGTCTATTATTGCTGAACAGCTTTTGCAGCAGCTTTTCAAGAAAGTCGTCCT 480  
DB |||||  
QY 7496 TTACTCAAGAAATTAATTAATGAACAGCTTTTCTGCAAGTACCAAGAGAGTCGTCCT 7555  
DB |||||  
QY 481 ATTTTAAACAGAGTTCAATTTGATTAATAGTAATCATAAAGATTTTAAAGCAGTAGGCACT 540  
DB |||||  
QY 7556 ATTTTAAACAGGTGCCACTTCTGATTAGTCAACACAAAGATTTAAACAGTTGCAACA 7615  
DB |||||



QY	121	ACAGGAGTA	CTTTAA	CAGGGT	CTAA	CGGT	CAAA	TATCA	AAAT	TCAA	ATTG	AAAA	CACA	CTATTCT	CGTA	180		
DB	7196	GAAGGT	ATTACT	TTTAA	TGGT	TTCA	AT	TGGT	CAAA	TGGT	CAAA	TTTCA	AT	TGGT	CAAA	7255		
QY	181	AGTAA	TGAAA	TGCT	GTGTT	GTCT	AA	TAC	CTCT	CC	CAG	GAG	CTA	TTTT	TA	TAG	AGCT	240
DB	7256	AAAA	TGAAG	ATGCT	GTGTT	GT	TTAA	TAC	TTCT	TTAG	GT	CGAT	CTCT	TTGA	AGCT	7315		
QY	241	TTTT	TATTA	TATTA	TTTCA	AGT	TTG	CC	CAG	A	TAT	TAG	TAT	AA	T	AA	T	300
DB	7316	TTCT	TAT	CA	AT	GT	AGT	AT	CT	AGT	TTA	CT	GA	TAA	CTCT	TTGA	TTA	7375
QY	301	CAAC	CA	AGT	TTG	TTTAA	CA	AGT	GTG	TA	AT	TAC	CT	TTAA	AG	GA	AA	360
DB	7376	CAAA	AT	CAAA	TG	TTTTA	AC	CA	GTG	GG	CA	AA	TAC	GA	AA	TTA	AG	7435
QY	361	GAC	CAG	TAT	CCT	CGT	CT	ACA	AG	AGT	AT	CA	CA	GAA	AA	TCT	CT	420
DB	7436	GA	CA	TAT	CCA	GA	AT	TCCA	AG	AA	TTT	CAG	CA	AG	CA	CT	CT	7495
QY	421	TTAT	TGA	GT	CTA	TTA	TG	CT	TGA	AC	AG	CT	TTT	TG	CA	GC	AG	480
DB	7496	TTACT	CA	AG	AA	AT	TAT	TAA	TGA	AC	AG	CT	TTT	TG	CT	CA	AG	7555
QY	481	ATTT	TAA	CA	GAG	TT	TC	ATA	TG	TAT	TAA	G	TA	AT	CA	TAA	AG	540
DB	7556	ATTT	TAA	CA	GAG	TT	TC	ATA	TG	TAT	TAA	G	TA	AT	CA	TAA	AG	7615
QY	541	GACT	CT	CAT	CGT	TAT	GAG	CC	AA	CGT	TTT	TA	AT	CAC	TTT	GG	CA	600
DB	7616	GACT	CT	CAT	CGC	CT	TA	AG	CA	AAAA	TT	GAC	CT	TT	CA	AAAA	AT	7675
QY	601	GTAG	TT	CTT	CCA	GT	AA	AT	CTT	TT	GAG	AG	AA	TTT	T	CAG	AG	660
DB	7676	GT	CGT	AA	TT	CT	TAG	CGG	AT	TTT	CAG	CGG	AT	TTT	TA	CA	AG	7735
QY	661	ACCG	TT	GA	GGT	AT	TTTT	CT	CA	CA	AG	CC	AA	AT	CTT	TT	G	720
DB	7736	ACT	GT	AG	AG	AT	TTCT	TT	GG	CA	AT	TA	CA	AA	AT	CTCT	TT	7795
QY	721	TAT	AC	AG	CC	CT	CTT	AGA	AG	AA	AT	TAT	CC	CG	AT	AC	AG	780
DB	7796	TAT	ACT	CGT	CT	CT	CT	AGA	AG	AA	AT	CT	CT	G	AT	CA	AG	7855
QY	781	GAGA	CG	GAG	GT	GT	TTTT	CA	AT	CC	CA	AT	CC	CT	CT	CG	CA	840
DB	7856	AACA	CT	ACT	AT	TACT	TTT	TA	TG	GT	TA	AA	CT	TAG	CC	AG	CT	7915
QY	841	ATTT	CT	TA	AT	GCT	ACT	CA	AAA	TG	GT	ACT	GT	TAA	GCT	TG	AG	900
DB	7916	TTAT	CA	AGT	CG	ACT	CA	AAA	TG	GT	ACT	GT	GAA	CT	TG	AA	AT	7975
QY	901	GCT	CA	TG	TT	AA	CT	CA	CT	CG	AG	TTG	GT	GT	AA	AG	GA	960
DB	7976	GCC	AT	GT	TT	CA	CT	CT	CA	GA	AG	TTG	GT	GT	AA	AG	GA	8035
QY	961	TCT	GGT	AGT	CA	TTT	AA	CT	AT	CAG	CTT	CA	AT	CC	CA	CTT	AC	1020
DB	8036	ACT	GGT	GAA	AG	AT	TGA	CC	AT	TAG	TTT	CA	AC	CC	AA	CTT	AT	8095
QY	1021	ATT	AAAA	AGT	GAAA	CA	GT	AAAA	AT	CT	TTT	TA	TC	CA	CAG	TT	CA	1080
DB	8096	TTAA	T	AG	CG	AA	AG	GT	G	ACT	AT	TAG	CT	TT	AT	CT	CAG	8155
QY	1081	CC	AG	CG	AT	GAG	G	AA	AG	AT	TTT	TAT	CC	AA	TT	TA	AT	1133
DB	8156	CC	AG	CA	G	AT	CT	G	CA	AG	AA	CT	T	C	AT	T	AC	8208
RESULT 24																		
BD003700																		
LOCUS																		
BD003700 21338 bp DNA linear PAT 31-JAN-2002																		

RESULT 24  
BD003700  
LOCUS

5

**DEFINITION** Polynucleotide of *Streptococcus pneumoniae* and sequence.

Qy	121	ACAGGAGTAACCTTTAAACAGGGCTAAACGGGTCAAAATATCAAAATGGAACACACTATTCCTGTGA	180
Db	7196	GAAGGTATTAATTTAAATTTGGTTCAAAATGGTCAAAATTTCAAAATGAAAAATTTATTTCTCTCAA	7255
Qy	181	AGTAATGAAAAATGCTGGTTTGCTAAATTTACCTCTCCAGGAGCTATTTTATTTAGAAAGCTAGT	240
Db	7256	AAAAATGAAGATGCTGGTTTGTAAATTTACCTTTTAGGTTCGATCCTCTTTGAAGCTTCT	7315
Qy	241	TTTTTTTAAATATATATTTCAAGTTTCCAGATATATAGTATAAATGTTTAAAGAAATTTGAA	300
Db	7316	TTCTTTATCAATGTAGTATCTAGTTTACCTGATGTAACTCTGTGATTTTAAAGAAATTTGAA	7375
Qy	301	CAACACCAAGTGTGTTTTAACCAAGTGGTAAATTCAGAGATTACCTTTAAAGAGAAAGATGTT	360
Db	7376	CAAAATCAAATTTGTTTTAAACAGTGGCAAAATCAGAAATTAACCTTAAAGAGAAAGATAGC	7435
Qy	361	GACAGATATCCTCGTCTCAAGAAGTATCAACAGAAAAATCCTTTGATTTTAAAAAAACAAA	420
Db	7436	GAACAAATATCCAGAAATCCAAGAAATTTACGACAGCACTCCTTTAACTTTGAAACAAA	7495
Qy	421	TTATTGAAGTCTATTATTTGCTGAAACAGCTTTTTCAGCCAGTTTACAGAAAGTCTGTCCT	480
Db	7496	TTACTCAAGAAAAATTAATAAGAAACAGCCTTTGCTGCAAGTACACAAGAGAGTCTGTCG	7555
Qy	481	ATTTTAAACGAGAGTTCATATTTGTATTAAAGTAATCATAAAGATTTTAAAGCAGTAGCGACT	540
Db	7556	ATTTTAAACAGGTGTCACCTTCGTATTGAGTCAACACAAAGAGTTTAAAAACAGTTGCAACA	7615
Qy	541	GACTCTCATCGTATGAGCCAAAGTTTAAATCATCTTTGGACAATTACTTCAGCAGATTTGATG	600
Db	7616	GACTCTCATCGCCTTAAGCCAGAAAAATTGACTCTTGAAAAAATAGTGATGATTTTGTAT	7675
Qy	601	GTAGTTCTTCCAAAGTAAATCTTTTGAGAGAAATTTTTCAGCAGTATTTACAGATGATATTGAG	660
Db	7676	GTCTGTAATTCCTAGCCGTTCTCTACGCCAAATTTTTCAGCGGTATTTACAGATGATATCGAA	7735
Qy	661	ACCGTTGAGGTAATTTTTCTCACCAAGCCAAATCTTGTTCAGAAGTGAACACATTTCTTTTT	720
Db	7736	ACTGTAGAGATTTCTTTTGCCAAATAACCAAAATCCTCTTTAGAAGCGAAAAATATTAGCTTC	7795
Qy	721	TATACAGCCTCTTAGAAGGAAATTTATCCGATACAGACCGTTTATTTAAATGACAGAAATTT	780
Db	7796	TATACTCGTCTCCTTAGAAGGAAACTATCCTGTATACAGATCGCTTTGATTCACACAGACTTT	7855
Qy	781	GAGACGGAGGTGTGTTTTCAATACCCAAATCCCTTCGCCACGCTATGGAACCGTCTTCCTTG	840
Db	7856	AACACTACTATTACTTTTAAATGTGGTAAACTTAGCCACAGTCAATGGAGCGTGCCCGTCTT	7915
Qy	841	ATTTCTAATGCTACTCAAAATGGTACTGTGTTAAGCTTGAGATTACTCAAAATCATATTTCA	900
Db	7916	TTATCAAGTCGACTCAAAATGGTACTGTGAAACTTGAATTAAGGATGGGGTGTGTAGC	7975
Qy	901	GCTCATGTTAACTCACCTGAGGTTGGTAAAGTAAACGAGGATTTAGATATTTGTTAGTCAG	960
Db	7976	GCCCATGTTCACTCCAGAAGTTGGTAAAGTAAACGAGAAATCGATCTGATCAGGTT	8035
Qy	961	TCGTGTAGTATTAATCACTCAGCTTCAATCCAACTTACCTTATGTAGTCTTTTAAAGACT	1020
Db	8036	ACTGGTGAAGATTTGACCAATTAGTTTCAACCCAACTTACTTGTATTGATTTCTCTTAAAGCT	8095
Qy	1021	ATTTAAAGTGAACAGTAAAAATTCATTTCTTATCACAGTTCGACCAATTCACCCCTAACA	1080
Db	8096	TTAAATAGCGAAAAGGTGACTATTAGCTTTATCTCAGCTGTTGCTGCAATTTACTCTTGTG	8155
Qy	1081	CCAGGGCATGAGGAAGAAAGTTTTTATCCAAATTAATTAACACAGTACCGAAACAAA	1133
Db	8156	CCAGCATACTGACGAAGACTTCATCGACTCATTAACAGGTTCTGTCAAA	8208

RESULT 24  
BD003700

DEFINITION	Polynucleotide of Streptococcus pneumoniae and sequence.
BD003700	
ACCESSION	BD003700.1 GI:18631661
VERSION	JP 2001501833-A/20.
KEYWORDS	unidentified
SOURCE	unidentified
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 21338)
AUTHORS	Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Baza,S.C., Fannon,M. and Dougherty,B.A.
TITLE	Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL	HUMAN: JP 2001501833-A 20 13-FEB-2001;
COMMENT	FATANT: GENOME SCIENCES INC OS Unidentified PN JP 2001501833-A/20 PD 13-FEB-2001 PF 30-OCT-1997 JP 1998520718 PR 31-OCT-1996 US 60/029960 PI CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI STEVEN C BARASH. PI MICHAEL FANNON BRIAN A DOUGHERTY PC C12N15/09,A01K67/027,C07K14/315,C07K16/12,C12N1/15,C12N1/19, PC C12N1/21, PC C12N5/10,C12P21/02.C12Q1/68,G06F17/30,C12N15/00,C12N5/00, PC G06F15/40 CC Strandedness: Double; CC Topology: Linear; FH Key Location/Qualifiers FT source 1..21338 FT /organism='Unidentified'.

FEATURES	Location/Qualifiers	FI	Organism	Unidentified

```

FEATURES
    source
        1. 21338
            /location/Qualifiers
            /organism="unidentified"
            /mol_type="genomic DNA"
            /db_xref="taxon:32644"

```

## ORIGIN

Query Match 53.6%; Score 608.2; DB 6; Length 21338;  
Best Local Similarity 71.1%; Pred. No. 3.2e-102;  
Matches 805; Conservative 0; Mismatches 328; Indels 0; Gaps 0;

1	ATGATTCAATTTTCAATTAATCGCACATTAATTATTTCATGCTTTAAATACAACCTAAACGT	60
7076	ATGATTCAATTTTCAATTAATAAAAATTTATTTCTACAAGCATTAATACTACTACAAGAGA	7135
61	GCTATTAGCACATAAAAAATGCGATTTCCTTATTTCTTTTCATCAATAAAAAATTGAAGTCACCTTCT	120
7136	GCTATTAGTCTTAATAATGCCATTTCCTATTTTATCAACAGTAAAAATTGACGTGACCAAT	7195
121	ACAGGAGTAACTTTAAACAGGGTCTAACGGTCAAATATCAATTTGAAAACACTATTTCCTGTGA	180
7196	GAAGGTATTACTTTAATTTCGTTCAAAATGGTCAAAATTTCAATTGAAAATTTTATTCTCAA	7255
181	AGTAATGAAAAATGCTGGTTTGCTAAATTACCTCTCCAGGAGCTATTTTATTAGAAAGCTAGT	240
7256	AAAAATGAAGATGCTGGTTTGTTAATTACTCTCTTAGGTTTCGATCCCTCTTGAAGCTTCT	7315
241	TTTTTTTAAATAATTAATTTCAAGTTTGGCCAGATATTAGTATAAATGTTAAAGAAATTGAA	300
7316	TTCTTTTATCAATGTAATGTAATCTAGTTTACCTGATGTAACTCTTGATTTTAAAGAAATTGAA	7375
301	CAACACCAAGTTGTTTTTAACCAAGTGTAATCAGAGATTACCTTTAAAGGAAAAGATGTT	360
7376	CAAAATCAAAATGTTTTTAACCAAGTGCGAAATCAGAAATTACCCCTAAAGGAAAAGATAGC	7435
361	GACCAGTATCCTCGTCTCAAGAAGATGATCAACAGAAAAATCCTTTTGATTTTAAAAAACAAAA	420
7436	GAACAATATCCAGGAATCCAGAAATTTGACGAAGCACTCTTTAATCTTGAACAAAA	7495
421	TTAATTGAAGTCTTAATATGCTGAAACAGCTTTTTCGACGCAGTTTTACAAGAAAGTCGTCTCT	480
7496	TTACTCAAGAAAAATTAATTAAGAAACAGCTTTTTCGTCGAAGTACACAAGAGAGTCGTCCG	7555



Qy	481	ATTTTAA	CAGGAGTTCATATTGTATTAAGTAATCATAAAGATTTTAAAGCAGTAGCGACT	540
Db	7556	ATTTTAA	CAGGTGCCACTTCGTATTGAGTCAACACAAAGAGTTAAAAACAGTTGGCAACA	7615
Qy	541	GACTCTCATCTGATGAGCAACGTTTAACTCACTTTGGACAATCTTCAGCAGATTTGATG	600	
Db	7616	GACTCTCATCCCTTAAGCCAGAAAAATTGACTCTTGAATAAATAAGTAGATGATTTTGAT	7675	
Qy	601	GTAGTTCTTCCAAGTAAATCTTTTCAGAGAAATTTTCAGCAGTATTTACAGATGATATTGAG	660	
Db	7676	GTGCTAAATCTTAGCCGTTCTCTACGCGAATTTTCAGCGGTATTTACAGATGATATCGAA	7735	
Qy	661	ACCGTTGAGGTATTTTCTCACCAAGCCAAATCTTGTTCAGAGTGAACAATTTCTTTT	720	
Db	7736	ACTGTAGAGATTTTCTTTGGCCATAACCAATCTCTTTAGAAGCGAAAAATATTAGCTTC	7795	
Qy	721	TATACACGCCCTCTTAGAAGGAAATTTATCCCGATACAGACCGTTTATTAAATGACAGAAATT	780	
Db	7796	TATACTCGTCTCTTAGAAGGAAACTATCTCTGATACAGATCGCTTGATTTCCAACAGAACTTT	7855	
Qy	781	GAGACGAGGTGTTTTCATATACCCAAATCCCTTCGCCACGCTATGGAACGTGCTCTCTTG	840	
Db	7856	AACTACTACTATTACTTTTAAATGTGGTAAACTTACGCCAGTCAATGGAGCGTGCCGTCIT	7915	
Qy	841	ATTCTTAATGCTACTCAAAATGTACTGTTAAAGCTTGAGATTACTCAAAATCATATTTCA	900	
Db	7916	TTATCAAGTCGACTCAAAATGCTATCTGTGAAACTTTGAAATTAAGATGGGTTGTTAGC	7975	
Qy	901	GCTCATGTTAACTCACTGAGGTTGGTAAAGTAAACGAGGATTTAGATATGTTAGTCAG	960	
Db	7976	GCCCATGTTCACTCTCCAGAAAGTTGGTAAAGTAAACGGAAGAAATCGATCAGGTT	8035	
Qy	961	TCGTGTAGTGATTTAACTACTACGTTTCAATCCAACTTACTTATTGAGTCTTTTAAAGCT	1020	
Db	8036	ACTGGTGAAGATTGACCAATTAGTTTCAACCCAACTTACTTGATTGATTTCTCTTAAAGCT	8095	
Qy	1021	ATTAAAAAGTGAAACAGTAAAAATTCATTTCTTATCACCAAGTTCGACATTTCAACCTTAA	1080	
Db	8096	TTAAATATGGCGAAAGGTGACTATTAGCTTTATCTCAGCTGTTTCGTCCTTACTCTTG	8155	
Qy	1081	CCAGGGCGATGAGGAGAAAGTTTATCCAAATTAATACACAGTAGCAACAAA	1133	
Db	8156	CCAGCAGATCTGAACGAGACTTCATGACGCTCATTTACACCAAGTTCGTACAAA	8208	

RESULT 25	SPNEU1901	232807 bp	DNA	linear	HTG 12-JUL-2001
LOCUS	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.				
DEFINITION	AL449923				
ACCESSION	GI:11545148				
VERSION	HTG; HTGS PHASE2.				
KEYWORDS	Streptococcus pneumoniae				
SOURCE	Streptococcus pneumoniae				
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.				
REFERENCE	1				
AUTHORS	Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de Francesco, M., Polissi, A., Buell, G., Feger, G., Garcia, E., Peitsch, M. and Garcia-Bustos, J.F.				
TITLE	Annotated draft genomic sequence from a Streptococcus pneumoniae type 19F clinical isolate				
JOURNAL	Microb. Drug Resist. 7 (2), 99-125 (2001)				
MEDLINE	2135329				
PUBMED	11442348				
REFERENCE	2 (bases 1 to 232807)				
AUTHORS	Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A., Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de Francesco, M., Buell, G., Feger, G., Garcia, E., Peitsch, M. and Garcia-Bustos, J.F.				
TITLE	Direct Submission				
JOURNAL	Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,				

COMMENT	FEATURES	source
Severo Ochoa 2, 28760 Tres Cantos, SPAIN		
* NOTE: This is a 'working draft' sequence. It currently		
* consists of 1 contigs. Gaps between the contigs		
* are represented as runs of N. The order of the pieces		
* is believed to be correct as given, however the sizes		
* of the gaps between them are based on estimates that have		
* provided by the submittor.		
* This sequence will be replaced		
* by the finished sequence as soon as it is available and		
* the accession number will be preserved.		
1 232807: contig of 232807 bp in length.		
Location/Qualifiers		
1. 232807		
/organism="Streptococcus pneumoniae"		
/mol_type="genomic DNA"		
/serotype="19F"		
/db_xref="taxon:1313"		
/clone="G54"		

Query Match	53.6%	Score 608.2	DB 2	Length 232807
Best Local Similarity	71.1%	Pred. No. 2e-103		
Matches 805	Conservative 0	Mismatches 328	Indels 0	Gaps 0
Qy 1	ATGATTCAAATTTTCAAATTAATCGCACATTAATTAATTCATGCTTTAAATAACAATAAAGCT	60		
Db 58105	ATGATTCAATTTTCAATTAATAAAAAATTAATTTCTACAAGCATTAATACTACTAAGAGA	58164		
Qy 61	GCTATTAGCACTAAAAATGCCATTCCTTATTCCTTCATCAATAAAAAATTCGAGTCATCTCT	120		
Db 58165	GCTATTAGTCTAAAAATGCCATTCCTATTTATCACAGTAAAAATTCGAGTCACCAAT	58224		
Qy 121	ACAGGAGTAATCTTTAAACAGGGTCTAACGGTCAAATATCAATTTGAAAAACACTATTCCTCTGA	180		
Db 58225	GAAGGTATTAATCTTTAAATGGTTCAAATGGTCAAATTTCAAATTTGAAAAATTTATTTCTCAA	58284		
Qy 181	AGTAATGAAAAATCGCTGGTTGCTTAATTAACCTCTCCAGGAGCTATTTTATTAAGAAGCTAGT	240		
Db 58285	AAAAATGAAGATCGCTGGTTGGTTGTAATTAATCTCTTTAGGTTCGATCCCTTTCTTGAAGCTTCT	58344		
Qy 241	TTTTTTTAAATAATTAATTTCAAAGTTTGCCAGAGATATTAGTATATAATGTTTAAAGAAATTCGAA	300		
Db 58345	TTCTTTATCAATGTAGTATCTAGTTTAACTGATGTAATCTTGATTTTAAAGAAATTCGAA	58404		
Qy 301	CAACACCAAGTTGTTTTAAACCAAGTGGTAAATCAGAGATTAACCTTAAAAAGGAAAAAGATGTT	360		
Db 58405	CAAAATCAAATTTGTTTTAAACCAAGTGGCAAAATCAGAAAAATACCTTAAAAAGGAAAAAGATGC	58464		
Qy 361	GACCAGTATCTCTGCTCAACGAAGTATCAACAGAAATTCCTTGATTTTAAACCAAAAA	420		
Db 58465	GAACAATATCCAGGAATCCAGAAATTTTCAGCAGCACTCTCTTTAAATCTTTGAACCAAAAA	58524		
Qy 421	TTAATCGAGTCTATTATTGCTGAAAACAGCTTTTGCCAGCCAGTTTCAAGAAGATCGCTCT	480		
Db 58525	TTACTCAAGAAAAATTAATAATGAAACAGCCTTTTGCTGCAAGTACACAAGAGATCGTCCG	58584		
Qy 481	ATTTTAAACAGGATTCATATTTGATTAATAGTAAATCAATAAGATTTTAAAGCAGTAGCGACT	540		
Db 58585	ATTTTAAACAGGATTCACACTTGATTTAGTCAACAACAAAGAGTTAAAAACAGTTGCAACA	58644		
Qy 541	GACTCTCATCGTATCAGGCCAACGGTTAATCACTTTGGACAATATCTTCAGCAGATTTTGATG	600		
Db 58645	GACTCTCATCGCTAAGCCAGAAAAAATTTGACTCTTGAAAAAATAGTGATGATTTTGAT	58704		
Qy 601	GTAGTTCTTCAAGTAAATCTTTGAGAGAAATTTTCAGCAGTATTTACAGATGATATTGAG	660		
Db 58705	GTCGTAATTCCTAGCCGTTCTCTACGGCAATTTTCAGCGGTATTTACAGATGATATCGAA	58764		
Qy 661	ACCGTTGAGGTATTTTTCTCAACCAAGCCAAATCTTGTTCAAGATGAAACATTTCTTTT	720		
Db 58765	ACTGTAGAGATTTCTTTTGCCAAATTAACCAATCTCTTTTGAAGAGCGAAAAATATTAGCTTC	58824		
Qy 721	TATACACGCCTCTTAGAAGGAAATTAATCCCGATACAGACCGGTTTTATTAATGACAGAATTT	780		

[illegible]

[illegible]

```
Db 4567 GCTATTAGTCTTAAAGTCCATTCCTATTATTTATCACTGTCAAAATTTGACGTGACCAAC 4626
Qy 121 ACAGAGTAGTAACTTTAACGGGTCTAACGGTCAAATATCAATTTGAAAAACATTTCTCTGTA 180
Db 4627 GAAGGTGTTACTTTAATGTTTCAATGCTCAAAATTTCAATTTGAAAAATTTTATTCTCAA 4686
Qy 181 AGTAATGAAGTCTGTTGTTGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 240
Db 4687 AAAAATGAAGTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 4746
Qy 241 TTTTATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 300
Db 4747 TTTTATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4806
Qy 301 CAACACCAAGTGTGTTTAAACCAAGTGTGTTTAAACCAAGTGTGTTTAAACCAAGTGTGTTTAA 360
Db 4807 CAAATCAAAATGTTTAAACCAAGTGTGTTTAAACCAAGTGTGTTTAAACCAAGTGTGTTTAAACCA 4866
Qy 361 GACCAATATCCAGTGTGTTTAAACCAAGTGTGTTTAAACCAAGTGTGTTTAAACCAAGTGTGTT 420
Db 4867 GAAATATCCAGTGTGTTTAAACCAAGTGTGTTTAAACCAAGTGTGTTTAAACCAAGTGTGTT 4926
Qy 421 TTAATGAAGTCTATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 480
Db 4927 TTAATGAAGTCTATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4986
Qy 481 ATTTTAAACAGGAGTTCATATGTTTAAACCAAGTGTGTTTAAACCAAGTGTGTTTAAACCAAGT 540
Db 4987 ATTTTAAACAGGAGTTCATATGTTTAAACCAAGTGTGTTTAAACCAAGTGTGTTTAAACCAAGT 5046
Qy 541 GACTCTCATGTTAGAGCAAGTTCATATGTTTAAACCAAGTGTGTTTAAACCAAGTGTGTTTAAAC 600
Db 5047 GACTCTCATGTTAGAGCAAGTTCATATGTTTAAACCAAGTGTGTTTAAACCAAGTGTGTTTAAAC 5106
Qy 601 GTAGTCTTCCAGTGTGTTTAAACCAAGTGTGTTTAAACCAAGTGTGTTTAAACCAAGTGTGTT 660
Db 5107 GTAGTCTTCCAGTGTGTTTAAACCAAGTGTGTTTAAACCAAGTGTGTTTAAACCAAGTGTGTT 5166
Qy 661 ACCGTTGAGGTTATTTTCTCACCAGGCAAGTTCATATGTTTAAACCAAGTGTGTTTAAACCAAGT 720
Db 5167 ACTGTAGAGATTTTCTCACCAGGCAAGTTCATATGTTTAAACCAAGTGTGTTTAAACCAAGT 5226
Qy 721 TATACAGCGCTCTTGAAGGAAATTTATCCCGATACAGACGTTTATTAATGACAGAAATTT 780
Db 5227 TATACAGCGCTCTTGAAGGAAATTTATCCCGATACAGACGTTTATTAATGACAGAAATTT 5286
Qy 781 GAGACGGAGTGTGTTTCAATACCAATCCCTTCGCGCATGATGAAAGTGTGTTTCAATACCAAT 840
Db 5287 AACCTACTATTACTTTTAAATGTTGTTTAAATGTTGTTTAAATGTTGTTTAAATGTTGTTTAAAT 5346
Qy 841 ATTTTCTAATGCTCAAAATGTTTAAATGTTGTTTAAATGTTGTTTAAATGTTGTTTAAATGTTTAA 900
Db 5347 TTAATCAAGTGTGCTCAAAATGTTTAAATGTTGTTTAAATGTTGTTTAAATGTTGTTTAAATGTT 5406
Qy 901 GCTCATGTTAACTCACTGAGGTTGTTTAAATGTTGTTTAAATGTTGTTTAAATGTTGTTTAAATGTT 960
Db 5407 GCTCATGTTAACTCACTGAGGTTGTTTAAATGTTGTTTAAATGTTGTTTAAATGTTGTTTAAATGTT 5466
Qy 961 TCTGTGAGTGTGTTTAAATGTTGTTTAAATGTTGTTTAAATGTTGTTTAAATGTTGTTTAAATGTT 1020
Db 5467 ACTGTGAGGATTTGACCAATGTTTAAATGTTGTTTAAATGTTGTTTAAATGTTGTTTAAATGTT 5526
Qy 1021 ATTTAAAGTGAACAGTGAATTTTAAATGTTGTTTAAATGTTGTTTAAATGTTGTTTAAATGTTTAA 1080
Db 5527 TTAATAGCGAAAGTGTGTTTAAATGTTGTTTAAATGTTGTTTAAATGTTGTTTAAATGTTTAAAT 5586
Qy 1081 CCAGCGGATGAGGAAAGTGTGTTTAAATGTTGTTTAAATGTTGTTTAAATGTTGTTTAAATGTTTAA 1133
Db 5587 CCAGCGGATGAGGAAAGTGTGTTTAAATGTTGTTTAAATGTTGTTTAAATGTTGTTTAAATGTTTAA 5639
```

```
AE008385 LOCUS 14222 bp DNA linear BCT 13-SEP-2001
DEFINITION Streptococcus pneumoniae R6 section 1 of 184 of the complete
genome.
ACCESSION AE008385
VERSION AE008385.1
KEYWORDS GI:15457528
SOURCE Streptococcus pneumoniae R6
ORGANISM Streptococcus pneumoniae R6
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 14222)
Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S.,
DeHoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C.,
Gilmour, R., Glass, J.S., Khoja, H., Kraft, A., Lagace, R.,
LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P.,
McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicas, T.I.,
Norris, F.H., O'Gara, M., Peery, R., Robertson, G.T., Rockey, P.,
Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G.,
Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L.
and Glass, J.I.
Genome of the bacterium Streptococcus pneumoniae strain R6
J. Bacteriol. 183 (19), 5709-5717 (2001)
21429245
11544234
2 (bases 1 to 14222)
Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S.,
DeHoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C.,
Gilmour, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R.,
LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P.,
McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicas, T.I.,
Norris, F.H., O'Gara, M., Peery, R., Robertson, G.T., Rockey, P.,
Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G.,
Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L.
and Glass, J.I.
Direct Submission
Submitted (27-JUL-2001) Infectious Diseases Research, Eli Lilly and
Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA
FEATURES
source
1. 14222
/organism="Streptococcus pneumoniae R6"
/mol_type="genomic DNA"
/strain="R6"
/db_xref="taxon:171101"
1. 1362
/gene="dnaA"
/note="synonym: spr0001"
1. 1362
/gene="dnaA"
/codon_start=1
/transl_table=11
/label=spr0001
/product="DNA biosynthesis, initiation, binding protein"
/protein_id="AAK98805.1"
/db_xref="GI:15457529"
/translation="MKKQFQNRILEFAQERLRSMDYFVIAQELIKVEENVATIFL
PRSEMYVWEKQDKIIVVAFGEIYDAEITPHYIFKPDQTSQVEATNLTLTDYS
PKLVISPYSDTKLEKVTDFNFIQDGNVAVSAALVSEDLATNLPLFYGGPGLG
KTHLLNAILGNILKNINARVYKIPAESFINDFLDHRLGEMKFKKTVRSLLDLLID
DIOSLGGKVAQTEFFENALHDKQKQIVLSDSPKHLGLEERLVTRSWGLTO
TITPPDETRIALQSKTEHLGYNQSDTLEYLAGQFNSVRDLGALINDITLIARVK
KIKDTIDIAAEAIRAKQDVQMLVPIPIQTEVGNFYGVSIKEMKSGRRQLQIVL
ARQVAMVLSRELTDNSLPKIGKFGGDKDHTVIIHAHAKIKSLIDQDDNLRLEIESIKK
KIK"
1521..2657
/gene="dnaN"
/note="synonym: spr0002"
1521..2657
/gene="dnaN"
/EC_number="2.7.7.7"
/codon_start=1
/transl_table=11
/label=spr0002
```







```
Qy 61 GCTATTAGCACTAAATGCAATTCCTATTCTTTTCATCAATAAATAAGTCACTTCT 120
Db 2368 GCTATTAGCACTAAATGCAATTCCTATTCTTTTCATCAATAAATAAGTCACTTCT 2427
Qy 121 ACAGGAGTAACCTTTAACAGGCTTAACGGTCAATATCAATGCAATATTCCTGTA 180
Db 2428 ACAGGAGTAACCTTTAACAGGCTTAACGGTCAATATCAATGCAATATTCCTGTA 2487
Qy 181 AGTAATGAAATGCTGTTGCTAATTAATACCTCTCCAGGAGCTATTTATTAGAAGCTAGT 240
Db 2488 AGTAATGAAATGCTGTTGCTAATTAATACCTCTCCAGGAGCTATTTATTAGAAGCTAGT 2547
Qy 241 TTTTATTATTAATATTATTCAGTTTCCAGATATTAGTATAAATGTTAAAGAAATGAA 300
Db 2548 TTTTATTATTAATATTATTCAGTTTCCAGATATTAGTATAAATGTTAAAGAAATGAA 2607
Qy 301 CAACACCAAGTGTGTTTAAACAGTGGTAAATCAGAGATTACCTTAAAGCAAGATGTT 360
Db 2608 CAACACCAAGTGTGTTTAAACAGTGGTAAATCAGAGATTACCTTAAAGCAAGATGTT 2667
Qy 361 GACAGATATCTGCTCTACAAGAGTATCAACAGAAATCTTTGATTTTAAACCAAAA 420
Db 2668 GACAGATATCTGCTCTACAAGAGTATCAACAGAAATCTTTGATTTTAAACCAAAA 2727
Qy 421 TTATTGAAGTCTATTATTCGTGAAACAGCTTTTGCAGCCAGTTTACAAGAAATGCTCT 480
Db 2728 TTATTGAAGTCTATTATTCGTGAAACAGCTTTTGCAGCCAGTTTACAAGAAATGCTCT 2787
Qy 481 ATTTTAACAGAGTCTCATATTGTTAATTAAGTAATCAATAAGATTTTAAACAGTAGGACT 540
Db 2788 ATTTTAACAGAGTCTCATATTGTTAATTAAGTAATCAATAAGATTTTAAACAGTAGGACT 2847
Qy 541 GACTCTCATGATGAGCAACCTTTAAT 569
Db 2848 GACTCTCATGATGAGCAACCTTTAAT 2876

RESULT 30
LOCUS AE006239 10186 bp DNA linear BCT 02-JUN-2004
DEFINITION Lactococcus lactis subsp. lactis IL1403 section 1 of 218 of the
complete genome.
ACCESSION AE006239 AE005176
VERSION AE006239.1 GI:12722835
KEYWORDS
SOURCE Lactococcus lactis subsp. lactis IL1403
ORGANISM Lactococcus lactis subsp. lactis IL1403
REFERENCE 1 (bases 1 to 10186)
AUTHORS Bolotin,A., Wincker,P., Mauger,S., Jaillon,O., Malarne,K.,
Weissenbach,J., Ehrlich,S.D. and Sorokin,A.
TITLE The complete genome sequence of the lactic acid bacterium
Lactococcus lactis subsp. lactis IL1403
JOURNAL Genome Res. 11 (5), 731-753 (2001)
MEDLINE 21235186
PUBMED 11337471
REFERENCE 2 (bases 1 to 10186)
AUTHORS Bolotin,A., Wincker,P., Mauger,S., Jaillon,O., Malarne,K.,
Weissenbach,J., Ehrlich,S.D. and Sorokin,A.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2001) INRA, Genetique Microbienne, Domaine de
Vilvert, Jouy en Josas 78352, France
FEATURES
source
1..10186
/organism="Lactococcus lactis subsp. lactis IL1403"
/mol_type="genomic DNA"
/strain="IL1403"
/sub_species="lactis"
/db_xref="taxon:272623"
358..1725
/gene="dnaA"
/feature="synonym: L0274"

gene
358..1725
/gene="dnaA"
/transl_table=1
/codon_start=1
/product="replication initiation protein DnaA"
/protein_id="AAK04099.1"
/db_xref="GI:12722836"
/transl_table="FASLNQKFWARVTELAQOISIGKQAYDFFIEPAQLMSVQDFTA
NILDSGMKDYWKQSLDITTAGFEVGRMIDYELVANDDELTELEHRLNQSIEE
QPRSTAKPAPLVSGLNEKNYFNPFVQPGNRTLAATAIADVPKPTDITPLFIYGG
GLKTHLMAICNOILTDNPTARIKYSSNFVNDYVNAATKNOMENFENYRLNLL
LLDDVDFPSDEKGTNEPNTFNALYDKGSOIULTSDRIPOELANLDRLSRFSMGL
TTDITAPDETREMAILLIKFSRSLRPESETLSYIAGQIDSNVRELSGALNRVFR
ANGTAVDIEETASQALSLKNATQOISNLIKIKIQDEVANYHISFSDLVGPKE
IAFPRQIAMLVIRELLGSLPAITGATFAGGRDHTTVMYAYQISDKMKNMDVQKDIDS
IKRKP"
1882..3024
/gene="dnaA"
/feature="synonym: L0275"
1882..3024
/gene="dnaA"
/feature="EVIDENCE BY HOMOLOG Y BIO10.02 DNA replication,
restriction, modification, recombination, and repair. DNA
polymerase III responsible for most of the replication
synthesis in bacteria. Exhibits also 3' to 5' exonuclease
activity."
/codon_start=1
/transl_table=1
/product="DNA polymerase III, beta chain (EC 2.7.7.7)"
/protein_id="AAK04100.1"
/db_xref="GI:12722837"
/transl_table="MIKFSIKNAFONALRITKQALGSKVTIPALTCLKLEVENGIT
LIGNGOISIKNPLPVDNKDASMLISCTGSLVLEAAFFENVSQLPTEPTEKEOK
QVLLTSGKSEITLKLDSLEYLPHLOEISGSSLMKVKVLEIFETFAVSTQENRP
IFGVHLEITLSTGELGADATDSHRMSQRLPDLSELSKVPILPSKINSFKNFTND
EEREIPIGSGMLFRNETIYSYRLEGSYPDNRILPNEADYTLDFVDAALRHT
MDARLLVTMTGTGKLVTSVSDSVTANSPEVSGVHEELTALSGENDLAISFNP
YLIDALKVIRAPVRIRFISNRPFTLQPRNEESGFVQLITPVRTN"
3130..6429
/gene="rexX"
/feature="synonym: L0252"
3130..6429
/gene="rexX"
/feature="EVIDENCE EXPERIMENTAL PMID:9435243 BIO10.01
Degratation of DNA. this exonuclease is specifically
blocked by the putative lactococcal Chi site
(5'-GGCGTG-3'). RexAB-mediated recombination stimulated
approximately 27-fold by lactococcal Chi."
/codon_start=1
/transl_table=1
/product="subunit B of ATP-dependent exonuclease"
/protein_id="AAK04101.1"
/db_xref="GI:12722838"
/transl_table="MEILTETQDITLGELEISLEELKRNKRVYIVPSSMFEKEK
EILRLAKGSDSAVFDLLVTRFKQPYFDKREKATTKTELGTAGLSMLFRVLRFS
KEEILYFLQDSAGLEMLIQLRTELNTANLSDNLPSPKQELKILSRFEKLA
NDYANSFEGDFTSLADGDFQKLDVTIVDGYTRFAEELFTESIQDVARFVI
GTYSNLSITAGSETIYISGOMIGFRSKFPVELKMAFSSVNVYVNLTKLLDLS
RFAISDNIEINSADAKYFRIWEAKQVEIEGVAKEIRKIQSGAFKFTVLGDDP
AAVEITLKEIFELYEIPFYAQESMSCHPLVIFPESLSIKKNYGTDDVNLKSK
VYTDVNLDEVIDYFYVQYKIKISGRKAFPAFSEFSEFKIELVNLQRENLLGNDSP
LQVFLGTRKQTKRWSDQLVLLNGVNMANNYTFSEASENKHQADHEQVQM
LISILNEFLAVFSDKELKSVFIDLILAGLKNKRYQIPANVNVNIQVLEYSK
KYIYALGSLQTFPRIKKNLTLSDDEERLEINQTDENQFIQLANVYVQKQFTVLS
LVNSAKETLVLMPOQIMANEQGFSPVFLNHSDEKTLQIKOENVLPSELEHNS
RSVISMIGIERELVETEKNDKRVFWSIIERILVKNPDQKILLDKAIDTVNL
SOETLDQIYGDKLVASVSSEFYVCEYQYFLETYLGLETPEINIDNLSNFFHVL
FEKMQEALSANDEKLTVLVDVDSYFETHTDATFTWNLSEIVQRTAVL
KETVSTDEKLTLLTSSFGPLSGELGNFSDVDIYLRGTRDLRDLSSDYLGAIDKSS
AHSFKLDQAYDGLSLQFMTYLDVIEAFPNQIKWALYQPKNQPTINLSIHLSEIA
GLLKESMYRVDGLVLEAADQIKAIENITVKKSNINQBEFQOLLKNENHYQAGRL
KSGQIATNIPMKRSEIGDQGNVRCRYCPKLSICRFEANVHMDHSRIGQKSQAEI
LAELKGBGRNE"
```



Qy	541	GACTCTCATGTAATGAGCGAAGCTTTAAATCACTTTGGACAAATACCTTTCAGACAGATTTGATG	600
Db	2422	GACTCATCGGATGAGTCAACGCTTTTATTGGCTTTTAAAGAGATTCAGAAATTTAAATTTGAT	2481
Qy	601	GTAAGTTCTTCCAGTAATACTCTTTGAGAGAAATTTTCAGCAGTATTTTACAGATGATATTGAG	660
Db	2482	GTTATTTTGGCAGTAATAATCAATAGATTTTAAAAATGTTTTTACAAATGATGAAGAA	2541
Qy	661	ACCGTTGAGGTATTTTTTCTCAACCAAGCCAAATCTTGTTTCAGAAAGTGAACACATTTCTTTTT	720
Db	2542	GAGATTTGAAATTTTCATAGTGGAGCCAAATGCTTTTATAGAAATGAACAAATAGTTAT	2601
Qy	721	TATACAGGCTCTTAGAAGAAATATCCCGATACAGACCGTTTATTAATATGACAGAAATTT	780
Db	2602	TATAGTCGTTTGAATTTGAAGGATCTTATCTCTGATACTAAATCGTTTAAATTCCAAATGAAGCT	2661
Qy	781	GA-----GACGGAGGTGTTTTTCAATACCANTCCCTTCGCCACGCTATGGAACGTCGCC	834
Db	2662	GATTAATCTTTAGATTTAGTTTTTTTGTATGCTGCGCAATTTGGCTATACAAATGGATCGTGCG	2721
Qy	835	TTCTTGATTTCTAATGCTACTCAAAATGGTACTGTTTAAAGCTTCGAGATTACTCAAAATCAT	894
Db	2722	CGTTTGTGACAGTAATGACAACTAATGGAACGTGTAATTAACGGTTTCTCGGTGATTTCT	2781
Qy	895	ATTTACAGCTCATGTTAACTCACCTGAGGTTGGTAAGGTAAACGAGGATTTAGATATTGTT	954
Db	2782	GTAGTAACACACAGCCAACTCTCTGAGGTAGGTTCTGTCATGAAGAAATTCAGCGCCTTA	2841
Qy	955	AGTCAGTCCTGGTAGTCAATTAACATATACGCTTCAATCCAACTTACCTTATTTAGTCTTTA	1014
Db	2842	TCAAAAGAGGAATGACTTAGCAATCAGCTTTAATCCCTGAAATTTTGTATGATGCTTTG	2901
Qy	1015	AAAGCTATTAAAAAGTGAACAGATAAAAATTCATTTCTTATCACCAGTTTCGACCATTCACC	1074
Db	2902	AAAGTGATTTAAAGCACCAGAGTTCGTATTCGCTTTATTTCTAATGTTTCGTCGTTCACT	2961
Qy	1075	CTAACACCGAGGCGATGAGGAAGAAAGTTTATCCAAATTAATACACAGTACGAACAAA	1133
Db	2962	TTACAGCCAGTAATGAAGAACTCGGATTTGTTCAATTGATTTACACCGAGTACGTACAAA	3020
RESULT	31		
LOCUS	LLU76424	10762 bp DNA linear	BCT 10-APR-1998
DEFINITION	Lactococcus lactis DnaA (dnaA) gene, partial cds; polymerase III subunit DnaN (dnan), exonuclease RexB (rexB), and exonuclease RexA (rexA) genes, complete cds; putative GTP binding protein gene, partial cds; and unknown gene.		
ACCESSION	U76424		
VERSION	U76424.1		
KEYWORDS	GI:2909712		
SOURCE	Lactococcus lactis		
ORGANISM	Lactococcus lactis		
REFERENCE	Bacteri; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.		
AUTHORS	1. (bases 1 to 10762)		
TITLE	el Karoui, M., Ehrlich, D. and Gruss, A.		
JOURNAL	Identification of the lactococcal exonuclease/recombinase and its modulation by the putative Chi sequence		
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 95 (2), 626-631 (1998)		
PUBMED	98118563		
REFERENCE	2. (bases 1 to 10762)		
AUTHORS	El-Karoui, M. and Gruss, A.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-OCT-1996) Genetique Microbienne, INRA, Domaine de Vilvert, Jouy-en-Josas 78356, France		
FEATURES	Location/Qualifiers		
source	1. .10762		
	/organism="Lactococcus lactis"		
	/mol_type="genomic DNA"		
	/strain="WG1363"		
	/db xref="taxon:1358"		

gene

1. .820

/gene="dnaa"

&lt;1. .820

/gene="dnaa"

/function="initiation of DNA replication"

/codon\_start=2

/transl\_table=11

/product="Dnaa"

/protein\_id="AAC12963.1"

/db\_xref="GI:2909713"

/translation="IKVSSNFVNDVYNAATRNQOMESFNTYRNLDLLDLDVQFFS  
 DKEGKNEFFNFALYDKGSQIVLTSRIPQELNNLEDLVSFRGWLTTDITADPY  
 ETRWAILLKSSEHSPSETLSYIAGQDSNVRELGALNVEFVARANGISIVDI  
 YTAQALRSKNATQOQSLNLTIKKIODEVANYTHSFSDLVGPKPKETAPPROJAM  
 YLRELCTSLPAICTAGCGRDHTTVWYAKQISDKMNDVQKDIDSIKKF"

977. .2119

/gene="dnaa"

977. .2119

/gene="dnaa"

/codon\_start=1

/transl\_table=11

/product="polymerase III subunit DnaN"

/protein\_id="AAC12964.1"

/db\_xref="GI:2909714"

/translation="MKFSINTAFQNALKITQATGSKVTIPALTKLIEVEERGIT  
 LIGSGOISIKNPLADNDASMLISGTGSVLEAFNFVVSQLPVLEFIEKEQK  
 QVLITGKSEITLKGDSIYIPLHLOISEGSSLMKMKVYLKEIFTEFVAVSTQENRP  
 FVHLEITLSTGELKAVATDSHRMSQRLPLEDTDLKFDVILPSKSINSKPNFTND  
 EEEIEIFISGSOQFQNETISYYSRLIEGYSPTNRLIENEADYTLDLVFDAAQRHT  
 MDRALRLVMTKINGTVKLTSGDSVVTANSPEVGSVEHELTALSKGENDLSISFNP  
 YLTDALVKVIAKPEVRIRFISNVRPFTLPQRNEESGFVQLITPVRTN"

2225. .5524

/gene="rexB"

2225. .5524

/gene="rexB"

/note="second subunit of major exonuclease"

/codon\_start=1

/transl\_table=11

/product="exonuclease RxB"

/protein\_id="AAC12965.1"

/db\_xref="GI:2909715"

/translation="MELIYTRITQDLTEGLLEIALELEKKNRVYIIVPSSPEKEK  
 EILERLAKGSDTAVDLVITREKTPQVYFDKREKATMTGLTGLSLMLKFRVLSFK  
 KDBIPLFLSDAGFLEMLIOLEAELTANLISVENPDNPNKELKMLKFLRAEELS  
 VEVANYSEGDFTNRLVDEPQDLKDVITIDGYRFSABEELFIESIQKVARFV  
 GYSDENSLTAGEIYVGTSMITFRKNFPVELRKAASAVNEVYSKLRILDLDS  
 RVVITDEKLKAEDEKFRIVEAENQVEIERVAKIRQIIQCAFPPKDFVLVGP  
 AAYEITLKEVFLYIEPIFYAQEESMSQHPVIFESLFAIKKNRYDDVNVLLKSK  
 VYTDANLDEEVDIYFYVQYKISGRKKFTFEESFESQIELVNMREKLGSSEP  
 LQVFLNNRKTGKKWSDLOGLLENGVMTNMNAYPSAELQNEHQWADKHEQVWM  
 LISTNEFLVPSDEKLSVRELDILLAGLNKAKYRQIPANVDVNVKDELVEPFTM  
 KYIYALGHSQNFPRPKNSLTLSDERLEINQITDENQFIEQLNVANYQKQFTVLS  
 LINSKESLVSMPQIMANQGEFSPVQLFKDADEKILQIKQVNLFSLEHIGNS  
 RSVTANTQIKRELVESEBEDKRVFWSIFRILVKSNDPQKILLDLAKDITVNL  
 APDTEUIGDKIYASVSESPERFVCEYQYPLENTLSLETENIDINSKLVGNFHEV  
 FEKVMETDLAENFDEKTLVLOEVKNYSRYFTODARTFWSNLEEIVROTATVL  
 KATVSDTLKTLITESSPLPKSELGNSVDVDIRGLIDELDLSLDYLGADYKGS  
 AHSFKLOAYDGLSLOFMYLDVIKQAPPNKKIYGLALYLOFKQPINLSINQUSIEA  
 NILKMEYEGVLIEDAEOIKGIENIALKKNINYNEEFELKLKNEHEHYRAAGQRL  
 KKGKAIPIPKRSEGIDQSGNVRCGRYCPUKSICRPFANIHMHNSREICQKQSAEI  
 LAELGEERDE"

5517. .9038

/gene="rexA"

5517. .9038

/gene="rexA"

/note="first subunit of major exonuclease"

/codon\_start=1

/transl\_table=11

/product="exonuclease RxB"

/protein\_id="AAC12966.1"

/db\_xref="GI:2909716"

/translation="MSEVKLTPEQNEAIHSSGKNILVSASAGSKTFVMAQRIVEKVK  
 QGIEIDLFIISTFTKKASELRLERLDLAKAQOESSDDDEAHLTLALQNLNSADIG

TWDSFTOKLTKANFNVRNIDPNFRILADQTESDLIROBVSQVLESYLSADESLNISK  
 DKFEKLIKNSKDRNIILOKVVVITYREASATENPISWLENQELKGETYKSLTDLIS  
 EDTVNVKENLLTREPILLEAISKKDPVYCTALFISIDTDIRVGSCKBALSKKDFLS  
 AQODLVGSKSPKRELKFKVDEKIKHQAFIEASDLQKQFIIDFYKTYLERK  
 KNEAFESDIAHFAIEILEENPDIRENLREHYDEIMIDEYQDTSHTOERMLELLSNG  
 HNLFWGDIKOSIYGFRLADPLGLFLEKYSDYDAENQNLIRLKENFRSGEVLNFTN  
 DIFRLMDKEKLGEMTYGKEALVQONTISDYPVEAKOPYPPELLYKENTSEETEDSE  
 VLSIDGEIKGAQBIKKIIEYGVPEKDIAILVRSNNKIEDILLISVDIPIVLDDEGR  
 VDFLKSMEVLMLDVRALDNLPLVDLSIVMLRSPLPQFNEDELTIRISVQGSRLRFW  
 DKILLKCKEKNPELNLISLEQKLKAFNQKFTWRKLVNKPPIHRLGLMKIYETYP  
 DYGALKNGENQANQALSVRAESYESSGYGLFKFVRLINKMEQNQNDLVNLIK  
 PQNAVTRMTPHKSGLFEDYVFLMNLQSFNRDLKEDVILSRHGLMKYIADLKE  
 DDVITDPYALVVMETPTFVWVKDLQKRAALSEEMRVLYAFTRAKKLYLVGKIKOT  
 FKKAGLELADTLEGLIKSDKFRNSRGFOHWILALQNAATKLPMKLVNVTOKLETE  
 KLEFTSPDFKALVESEKFDNISFSDIEKEAKOINWYQYPHQAATLSSIOPTPSQV  
 KKRSEYKQLOVGEVQVSEFVRVKNLDFSDGPKKITAENGSAHSFMOYADPFSQD  
 LFSQATLIDEMGDEKIKNOIDITKILITLFDTEFGQFLSENVDKTVKEAPFSMLRTDE  
 FAKEQYIVRGICDGFGLKADKILIDFYKTRFTNVSAISSEIKERYKQDMLYSEALQK  
 AYHVQIDKILILILGGRKVFVEKIDD"

/complement(9281..9838)

/note="putative ORF 5"

/codon\_start=1

/transl\_table=11

/product="unknown"

/protein\_id="AAC12967.1"

/db\_xref="GI:2909717"

/translation="MLGENIQKARLSKNLTQEBVAKELYFSROAISRWESNKTENPFE  
 TLALADLYSDSLIAFSQGIPEQKRKKHINLPAFGVLVFNFVLGFWLLIITALLLLA  
 LYGLILAIPLAPITMIPALVTHNPHTFGNANGATGDMWYWLVAIISCSLMVIVSFF  
 IWRFTKLLYFWLLKYLKFNLSVYN"

10122. .&gt;10762

/note="ORF 6"

/codon\_start=1

/transl\_table=11

/product="putative GTP binding protein"

/protein\_id="AAC12968.1"

/db\_xref="GI:2909718"

/translation="MALTAGIVGLNVGKSTLPNAITKAGAEANYPFATIDPNVGMV  
 EVDPERLNKLTILKPKKTVPTTFEFTDIAGIVKASRGEGKGNKFLANIREVDIAIH  
 VVRAFDENVMRENREDAPIDPMADIEINLELILADLESVYKRYARVEKARTAKD  
 KDAVAFNVLKLLKPVLEDSGARTIDFDEDETKVLKSLFLLTSKPVLYVANVS"

ORIGIN

Query Match

Best Local Similarity 36.0%; Score 408.2; DB 1; Length 10762;

Matches 690; Conservative 0; Mismatches 443; Indels 6; Gaps 1;

Qy 1 ATGATTCAATTTTCAATTAATCGCATTATTTATTTATTCATGCTTTAAATACAACTAAACCT 60

Db 977 ATGATTAAATTTTCAATTAATAAATGCGCTTTCAAAATGCTTTGAAATATTACCAACAA 1036

Qy 61 GCTATTAGCACTAAATAATGCCATTCTCTATCTTTTCATCAATAAAAATTGAAGCTACTTCT 120

Db 1037 GCATTTGGTTCTAAGTTCAATTCCTGCTTTTAAACAAAGTTTAAAAATTTGAAGTTGAAGAA 1096

Qy 121 ACAGGAGTAACTTTAAACAGGGTCTAACCGTCAAAATCAATCAATGAAACAACTATTCCTGTA 180

Db 1097 AAAGGAAATACCTTAATTTGGTTTCAATGGTCAATTTCAATCAAAAACCTTTTTCAGCT 1156

Qy 181 AGTAATGAAAGTCTGGTTTGTCTAATTTACCTCTCAGGAGCTATTTTATTTAGAAGCTAGT 240

Db 1157 GATAATAAAGATGCAAGTATGCTTTATTTTTCAGAACTGGTCTCTGCTTATTAGAAGCTGCT 1216

Qy 241 TTTTATTATTAATTTATTTTCAAGTTTTCAGATATTAGTATAAATGTTTAAAGAAATTGAA 300

Db 1217 TTCTTTGAAAGTAGTAGTAACTCACTTCTGAGTTACTTTGGAAATTTATCGAAAAAGAA 1276

Qy 301 CAACACCAAGTTGTTTTAACCGTGGTAAATCAGAGATTACCTTTAAAGAGAAAGATGTT 360

Db 1277 CAAAAACAAGTTTGTGTTGACTTCTGGTAAATCAGAAATTAATCTTTAAAGAGTTAGATTCT 1336

Qy 361 GACCAGTATCTCGTCTACAAAGAGTATCAACAGAAATCTCTTGAATTTTAAAAACAAA 420

Db 1337 GAAATTTATCCACATTTCAAGAAATCTCTGAAGGTTCTTCTTTAAATGAAGTGAAA 1396

QY 421 TTATTGAAGCTATTATTGCTGAAACAGAGCTTTTGGAGCAGGTTTCAAGAAAGTCTGCT 480

Db 1397 GTTTTGAAGAAATTTTACAGAGAGCTGTTTTTGTCTGAAGTACTCAAGAAATGCTCCA 1456

QY 481 ATTTTAAACAGAGGATTCATATTTCTATTAGTAAATCATATAAGATTTTAAAGCAGTACGACT 540

Db 1457 ATTTTCAACAGGGGTTTCACTTGAATCTTTGTCAAGGGTGAATTTAAAGCAGTACGACT 1516

QY 541 GACTCTCTGATGATGAGCAACGTTTAAATCACTTTGGCAATVACTTTCAGCAGATTTTGATG 600

Db 1517 GACTCAGATCGAATGAGCAACGTTTATGCTTTTAGAGATACAGACTTTAAATTTGAT 1576

QY 601 GTAGTTCTTCCAGTAATCTTTGAGAGAAATTTTTCAGCAGATTTTACAGATGATTTGAG 660

Db 1577 GTAAATTTTACCAAGTAATCAATCAATAGTATTTTAAATATGTTTTTACAAATGATGAAGA 1636

QY 661 ACCGTTGAGGATTTTCTCACCAGCCAAATCTTGTTCAGAGTGAACACATTTCTTTT 720

Db 1637 GAGATTGAATTTTCACTAGTGAAGCCAAATGCTTTTCCAAATGAACAAATTAGTTAT 1696

QY 721 TATACACGCTCTTGAAGGAAATATCCCGATACAGACCGTTTATTAATGACAGAAATTT 780

Db 1697 TATAGTCTGTTGATGAGGGTCTTATCTCTGATCTATCTGTTGATTCCTCAATGAGCT 1756

QY 781 GA-----GACGAGGTTGTTTCAATACCCAAATCCCTTCGCCAGCTATGGAACGTGCC 834

Db 1757 GATTATCTTTAGATTGCTGTTTGTGCTGCGCAATTCGCTCATCAATGATCGTGCA 1816

QY 835 TTCCTGATTTCTAAGTCTACTCAAAATGCTACTGTTTAACTTGAGTTTCAATCAATCAT 894

Db 1817 CGTTTATTGACAGTAAATCAACCAACGGAACAGTAAATGACTGTTTCAGGTGATCT 1876

QY 895 ATTTTCAGCTCATGTTAACTCAGTCTGAGTGTGTTAGGTAAGCAAGGATTTAGATATTGT 954

Db 1877 GTCTGAACACAGCAAACTCTCTGAGTAGTGTCTGTTCAATGAAGACTGACAGCTTA 1936

QY 955 AGTCAGTCTGTGATGATTAATCAATCACTGCTTCAATCCAACTTACCTTTATGAGCTTTTA 1014

Db 1937 TCAAAAGAGGTAATGATTTATCTATCATCTGCTTCAATCTGAGTATTTGATGATGCTTTG 1996

QY 1015 AAAGCTATTAAAGTGAACAGTAAATTTCTATCTATCCAGTTCAGCTTGCACATTCACC 1074

Db 1997 AAAGTGAATTAAGCACCAGAGTACGAATTCGTTTCATTTCAACGTTTCGTCGCTCACT 2056

QY 1075 CTAACACACAGGCGATGAGGAAGAAAGTTTTATCCAAATTAATTAACACAGTACGAACAA 1133

Db 2057 TTGCAACACGGAATGAAGAACTGGAATTTGTTCAATGATTACACAGTACGATCAAA 2115

RESULT 32  
AE016947

LOCUS AE016947

DEFINITION Enterococcus faecalis V583, section 1 of 11 of the complete genome.

ACCESSION AE016947

VERSION AE016947.1

KEYWORDS GI:29342119

SOURCE Enterococcus faecalis V583

ORGANISM Enterococcus faecalis V583

REFERENCE 1 (bases 1 to 326434)

AUTHORS Paulsen, I., Banerjee, L., Myers, G.S.A., Nelson, K.E., Seshadri, R., Read, T.D., Fouts, D.E., Eisen, J.A., Gill, S.R., Heidelberg, J.F., Tettelin, H., Dodson, R.J., Umayam, L., Brinkac, L., Beanan, M., Dougherty, S., DeBoy, R.T., Durkin, S., Kolonay, J., Madupu, R., Nelson, W., Vamathevan, J., Tran, B., Upton, J., Hansen, T., Shetty, J., Khouri, H., Utterback, T., Radune, D., Ketchum, K.A., Dougherty, B.A., and Fraser, C.M.

TITLE Role of Mobile DNA in the Evolution of Vancomycin-Resistant Enterococcus faecalis

JOURNAL Science 299 (5615), 2071-2074 (2003)

PUBMED REFERENCE AUTHORS

TITLE JOURNAL

FEATURES source

gene

CDS

gene

CDS

gene

CDS

gene

CDS

12663927

2 (bases 1 to 326434)

Paulsen, I., Banerjee, L., Myers, G.S.A., Nelson, K.E., Seshadri, R., Read, T.D., Fouts, D.E., Eisen, J.A., Gill, S.R., Heidelberg, J.F., Tettelin, H., Dodson, R.J., Umayam, L., Brinkac, L., Beanan, M., Dougherty, S., DeBoy, R.T., Durkin, S., Kolonay, J., Madupu, R., Nelson, W., Vamathevan, J., Tran, B., Upton, J., Hansen, T., Shetty, J., Khouri, H., Utterback, T., Radune, D., Ketchum, K.A., Dougherty, B.A., and Fraser, C.M.

Direct Submission

Submitted (03-FEB-2003) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

Location/Qualifiers

1.326434

/organism="Enterococcus faecalis V583"

/mol\_type="genomic DNA"

/strain="V583"

/db\_xref="taxon:226185"

59..1402

/gene="dnaA"

/locus\_tag="EF0001"

59..1402

/gene="dnaA"

/locus\_tag="EF0001"

/note="identified by match to PFAM protein family HMM PF01695"

/codon\_start=1

/transl\_table=11

/product="chromosomal replication initiator protein DnaA"

/protein\_id="AAO79886.1"

/db\_xref="GI:29342120"

/translation="MPDVEFWHSLEAYQALSPFDMWIKTRPLKLDNNQWLVE VPSAHRDWEKNLSAKIVETFKLTGAEMPHFVVADEKAAALAELEPAAEEVVF SEQKKAMLPKTYTDFEVIQKGMQMAAALVAEDFGSIYNPLFFGGVGLKTHL MHATQHOMLPDAKVKVSVETFTNEFINSTQTKSEQRKVRNVDLLVDIQQF LAEATLEEFPHFTNDLYNENKQIVLTSDDPNDI PKLPERLVRFAWGLSVDTIPP DLEFRIALRKADAARLEIPDDTSLYIAGDINSIRELEGALVRVQAPATNGEDIT TSLAADKLSKSVGSKNQSLIQOESVKYHVPVLDKLGKGRVKIIVPRQISM LAREMTDNLSPKIGAEFGGKHDTTVIHAHEKIQLLEKDPALQKEVSEIKNLLNS "

1604..2734

/gene="dnaA"

/locus\_tag="EF0002"

1604..2734

/gene="dnaA"

/locus\_tag="EF0002"

/note="similar to GB:X69151, GB:J05682, SP:P21283, PID:340188, and PID:37643; identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=11

/product="DNA polymerase III, beta subunit"

/protein\_id="AAO79887.1"

/db\_xref="GI:29342121"

/translation="MKLTVRKSVFLQELQTVQRAISSKTTTIPILTVKVLSEDLGL TGSNADISIESFLSKDDKAQMTERTQSIVLQSFEGEIIKLPEDMTMELVNNQ VAITGKADFTVGLDADNYPHLPVIDTQNMKQFVHLTKIISTGFAVSHSRPI LTVGHFIDENQKLLAVATDSHRLSORVPTQEAQVEDFNIVPGKLSLSTNEEBE NVEISIMENQVLFKTYMYSRLLGNYPDTPNLTSHNTQIEFYFPELLSAIERA SLLSGHGNNTVRLSISPDVLYGNSPEIGKVEALNENVSGEALDISFNPYMKD ALRAFQDNNTVKLSPIRPTLEPTETELDIQLITPVRTN"

3313..3558

/locus\_tag="EF0003"

3313..3558

/locus\_tag="EF0003"

/note="similar to SP:P05650; identified by sequence similarity; putative"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAO79888.1"

/db\_xref="GI:29342122"

/translation="MKKTFVLATEVITIGQFLKEINVIGSGGQAKWYLNADNSVFDVDE LENRGRKLYAGMMIEIEEGTFFFMVKNESDDAAE"

```

gene
3545. .4672
/genes="recF"
/locus tag="EF0004"
3545. .4672
/genes="recF"
/locus tag="EF0004"
/notes="identified by match to PFAM protein family HMM
PF02463"
/codon start=1
/transl table=11
/product="DNA replication and repair protein RecF"
/protein_id="AAO79889.1"
/db_xref="GI:29342123"
/translation="MRLNELTQHYRNYETVSLDFPKTLNLFGENAOGKTNLLRSIY
VLAMTRSHRTSNEKELIGWEAQAKISGVVEKKTGTVPLEILI SNGKRTKYNHLEOK
RLSAYIGQLNVLFAPELDSLAVKGSPOVRERKIDMELGQSPILYLDLYOVSVLKQR
NOYLKOLAEKOTDTVYLDLLEQALAEFGGKVLVYARLGLKLEHWANLHGHKSHGR
EYLTDYASISPIDNTDLSLEALONQLQOLMNNRRKELFKANTFLGPHRDDLLFIVN
GNQVYIGSQGQRTALSILKLAIEDLMHSETGEYFVLLDDVMSBELDNERQIHLEET
IEKVQTFUTTSLDHIDKLTVEPDI FYVQOQKIERNAT"
4690. .6618
/genes="gyrB"
/locus tag="EF0005"
4690. .6618
/genes="gyrB"
/locus tag="EF0005"
/notes="similar to GB:D16626, SP:P42357, and PID:451210;
identified by sequence similarity; putative"
/codon start=1
/transl table=11
/product="DNA Gyrase, B subunit"
/protein_id="AAO79890.1"
/db_xref="GI:29342124"
/translation="MREAEQYDASQIOVLEGLAEVRKRPYGIYSTGEGHLLVWE
IVDNSIDEALGAFKSIQVIEPDSITVIDGRGIPVGIOAKTGRPAVEVTFVLA
GKGFGGGYKVGSGHGGSSVVALSTSLDVYVVDKGYQYQYRGAUVDDLLKVEI
ETDRGTTVHFIPDEIFETIYVDFDLKATRLAFNLRLGLHLSIEDRREGQEDKK
EYHGGGKSYVHNLNANDVIFPEPIEGEQDITVEVSQYTDGYHSLNLFSPAN
IHTYGGTHSGESGKSLTRVINDYARKOKLMKDEKLTGEDVREGLTAVVSIKHPD
QPEGKTKLGNSEYVTRDLRFSEYFTFKMENPTVGKQIVKGMASKAKLAAKRA
REVTRKGALEISNLPKGLADCSKDPKCELFIVEGDSAGSAGKQGRSRFOALPI
RGIILNVEKASMDKILANEEIRSLFTAMGTGEGEDFDVSKARYKHLVMTDADVDGAH
ITLLTLFYRPMRPIVEAGYVYIAQPLYGKQKNITYVQPGHAESEBLKLEEL
PASFPQVQRYKLGEMDQHLWETMPDPEKRLMARVSVDDAIEADQIFEMLMGDRVE
PRAFIENAHYVKNLDI"
6674. .9175
/genes="gyrA"
/locus tag="EF0006"
6674. .9175
/genes="gyrA"
/locus tag="EF0006"
/notes="similar to GP:7532887; identified by sequence
similarity; putative"
/codon start=1
/transl table=11
/product="DNA Gyrase, A subunit"
/protein_id="AAO79891.1"
/db_xref="GI:29342125"
/translation="MSEIEKENIQDNLITSEMKEFIDYAMSVIVARALPDVRDLKP
VHRRILYGNELGVTPDPKPKKSARIVGDVNGKYPHGDSDAIYESVMYMAQPFYSVRAM
LVDHGNFGSDVGDCAAMRYTEARMSKIALEMLRDINKNTVDVFGQYVDSEQBPVL
PARPNLVNCTGTGAVGWATNIPHLNISEVIDATSLMDPNVDVTHLMEVLPFGPD
PTGLVMGKSGIRRAYETGKGIIVRAKVELTEMPNGKERILVTELPVWNAKILIER
ISLHRDKRIGITDLRDESSREGRIVIDVRDVSASVNLNLYKMTALQTSFGFNM
LAIEKGVKILSLKRLNLYVEHQEVITRETFIDKNKAEARAHILEGLRIALDHI DE
IIAIRGSDDEAKATILIERFESDQAQAILDMLRLRLTGLERDKTENYQBLKF
IADLEDILARPERVETIITBLNDVRTKFGDARTELLVGLVLEDEDLIEEVEVIT
TLNNGYIKPMANSEFRQRRGQVGVGHDDDFVKNLVSCSTHTLFLFFNTGKV
YRAGYELPYGRYKAGIPVNLIGIDSAEKIQALISVEGKAEAGKVLFFTLKGTVK
RTATVAFNIAISGALISLAKDEDELVDVNTNGNQMIIIGHAGYSVTFDFTLKVYK
GRITAGVGNPLRENDYVVGAAIILDENKENVLITENGYGRKTKASEYFVKRGKGKIK
TANTKNGPLAGLTVDGDEDIILITNKGVIIRFNVDVSVSQTGRATLGLVRLMEDG
AKVVTMAVVEPEVEEIVEVETNETTSETEE"

```

```

gene
9350. .9652
/genes="rpsF"
/locus tag="EF0007"
9350. .9652
/genes="rpsF"
/locus tag="EF0007"
/notes="identified by match to TIGR protein family HMM
TIGR00166"
/codon start=1
/transl table=11
/product="ribosomal protein S6"
/protein_id="AAO79892.1"
/db_xref="GI:29342126"
/translation="MSQTKYEIMVIRPNIDEAKTALVERFDITLKDNGAEVIESK
DWEKRLAYKNGFRREGIYHIVNVTSPSTAGINEFDRKLAKINDDIIRHMVKVEA"
9697. .10263
/genes="rpsB"
/locus tag="EF0008"
9697. .10263
/genes="rpsB"
/locus tag="EF0008"
/notes="similar to GP:9789553, and GP:9789553; identified
by sequence similarity; putative"
/codon start=1
/transl table=11
/product="single-strand binding protein"
/protein_id="AAO79893.1"
/db_xref="GI:29342127"
/translation="MINNVLVRLTKDPLRYTASGASAVATFTLAVNRNFTNQDGR
EADFINCVIRKPAETWANYARKTLLGVGRITRNVYQQQGVVTVVCENFOL
LESRSASEQRTGGGSGFNNGYQSNRSGFNNGYQSNRSGFNNGYQSNRSGFNNGY
GMPDFDKSDPFGGSGSIDISDDLLPF"
10289. .10528
/genes="rpsR"
Query Match 29.9%; Score 339.4; DB 1; Length 326434;
Best Local Similarity 56.8%; Pred. No. 4.5e-53;
Matches 644; Conservative 0; Mismatches 486; Indels 3; Gaps 1;
Qy 1 ATGATTCATTTTCAATTAATCGCACATTTATTTATTCATGCTTTTAAATACAACTAAACGT 60
Db 1601 ATTAGTAATTAACGGTCAACGAAGTGTCTTTTACAAGAAATTTTCAAACTGTTTCAACGA 1660
Qy 61 GCTATTAGCACTAAAAATGCCATTCCTATTCTTTTCAATCAATAAAAAATTTGAAGTCACTTCT 120
Db 1661 GCAATTTCTTCTAAAACACGATTCCTAATCTTAACAGGTGTAAATTTGTGTTATCAGAA 1720
Qy 121 ACAGAGTAATCTTAAACAGGTCCTAACGGTCAAAATATCAATTTGAAAAACACTATTCTCTGTA 180
Db 1721 GATGGCTTATCACTTACTGGGAGTAACGGGATATTTCAATTTGAAAGTTTTTTTAAAGTAAA 1780
Qy 181 AGTAATGAAATGCTGTTGCTTAATTAACCTCTCCAGAGCTATTTTATTAGAGAGTAGT 240
Db 1781 GATGATGAAAGCTCAATGACCAITGAAACGACAGTAGCTGTTTACAAATCTCGT 1840
Qy 241 TTTTATTATTAATTTTCAAGTTTCCAGATATTAGTATAAATGTTTAAAGAAATTTGAA 300
Db 1841 TTTCTTGGCGAAATTTATTCGTAATTAACAGAGATATGTTCAATTTGGAAGTTCTAGAT 1900
Qy 301 CAACACCAAGTTGTTTAAACAGTGGTAATTCAGAGATTACCTTTAAAGGAAAAAGATGTT 360
Db 1901 AACAAATCAAGTTGCAATTTACTTCAGGAAAAAGCTGATTTTACGGTTTATGGAATTAGATGCT 1960
Qy 361 GACCAGTATCTCTGTCTACAAGAGATATCAACAGAAATCTTTGATTTTAAAAACAAA 420
Db 1961 GATAATTTATCGGACTTACAGTAATGATACTCAAAACCAATGAATTAACCTGTTCTAT 2020
Qy 421 TTAATGAAGTCTATTATGCTGAAACAGCTTTTTCAGCCAGCTTTTACAAGAAAGTCGTCCT 480
Db 2021 TTAATGACAAAAATTTATTAGTAAACAGGTTTGTGCTATCGATGACGAAAGTCGTCCT 2080
Qy 481 ATTTTACAGAGTTCATATTGATTAGTAAGTAATCAATAAGATTTTAAAGCAGTAGGCACT 540

```

Db 2081 ATTTTAACTGGGTTCACTTTATTTTATTTAGAAATCAAAA---TTACTTGGCGTTGGACA 2137  
QY 541 GACTCTCATCGPATGAGCAACGTTTAACTTTTGGACAAATCTTCACAGAGATTGTATG 600  
Db 2138 GATTACATCGTTTAACTCAAGTGTATCCGACAGAACAGCAGTAGAGACITTTAA 2197  
QY 601 CTAGTTCTTCCAGTAATCTTTGAGAGATTTTTCAGCAGTATTTTACAGATGATATTGAG 660  
Db 2198 ATTGTAAATTCAGGAAAAAGTTTAACTGAACCTTTCTCGTTCAATTAACCAATGAAGAAA 2257  
QY 661 ACCGTTGAGGTATTTTCTCACCAAGCCAAATCTTTGTTTTCAGAAAGTGAACACATTTCTTTT 720  
Db 2258 ATGTTTGAATCAGCATTTATGAAATCAAGTCTATTTTAAACAGAAACATGTACTTC 2317  
QY 721 TATACACGCTCTTGAAGAAATTTATCCCGATACAGACCGTTTATTAATGACAGAAATTT 780  
Db 2318 TATTCTCGTTTGTAGAGAAATTTATCTGTATACCAACCGTTTAACTCCAACTAGCCAT 2377  
QY 781 GAGACGAGGTGTTTTCATACCAATCCCTTCCGACGCTATGGAAGGTGCTTCTTG 840  
Db 2378 AACGCAAAATGAAATTTTATGTACCAAAATTTGTTTTCAGCAATCGAAGTGCCTCTTTA 2437  
QY 841 ATTTCTAATGCTACTCAAAATGCTACTGTTAAGCTTTGAGATTACTCAAAATCATATTTCA 900  
Db 2438 CTTTCTCATGAGGAGGTAAACATATTTGTCCTTTCAATTTACAGATTCGTGTT 2497  
QY 901 GCTCATGTTAACTACCTGAGGTTGTTAAGTAAACGAGATTTAGATATTTAGTCAG 960  
Db 2498 TTATATGAAATTCACCTGAAATTCGAAAGTGAAGAGCTTTAACTATGAAATGTT 2557  
QY 961 TCTGTAGTATTTAACTATCAGCTTCAATCAACTTACCTTATTTGAGTCTTTAAAGCT 1020  
Db 2558 TCTGTGAGCGTTGGATATTTCTTTCAACCCAGATTTATGAAAGATCGGTTGCGAGCG 2617  
QY 1021 ATTTAAAGTGAACAGTAGTAAATTTCAATTTCTTATCACAGTTTCGACCATTCACCCCTAACA 1080  
Db 2618 TTGCGGATGAAATTTACCGTGAATTCCTTTCTCCAAATTCGTCGTTACATTTGAG 2677  
QY 1081 CCAGCGATGAGGAAGAAGTTTATCCAAATTTATACACAGTACGAAACAA 1133  
Db 2678 CCAACCGAACAGAACTAGATTTCATTTCACTAAATTTACACCGTTTCGTACAA 2730

## RESULT 33

LOCUS BD194028 8001 bp DNA linear PAT 17-JUL-2003  
DEFINITION Enterococcus faecialis polynucleotides and polypeptides.  
ACCESSION BD194028  
VERSION BD194028.1 GI:33003767  
KEYWORDS JP 2002529046-A/622.  
SOURCE unidentified  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 8001)  
AUTHORS Kunsch,C.A., Dillon,P.J. and Barash,S.C.  
TITLE Enterococcus faecialis polynucleotides and polypeptides  
JOURNAL Patent: JP 2002529046-A 622 03-SEP-2002;  
HUMAN GENOME SCIENCES INC  
COMMENT OS Unidentified  
PN JP 2002529046-A/622  
PD 03-SEP-2002  
PF 04-MAY-1998 JP 1998548302  
PR 06-MAY-1997 US 60/044031.16-MAY-1997 US 60/046655 PR  
PI CHARLES A KUNSCH,PATRICK J DILLON,STEVEN C BARASH PC  
C12N15/31,C07K14/315,C07K16/12,C12Q1/68  
CC Strandedness: Double;  
CC Topology: Linear;  
CC Enterococcus faecialis polynucleotides and polypeptides FH  
Key Location/Qualifiers  
FT source 1..8001  
FT Location/Qualifiers /organism='Unidentified'.  
FEATURES

source 1..8001  
/organism='unidentified'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32644'  
ORIGIN  
Query Match 29.2%; Score 330.6; DB 6; Length 8001;  
Best Local Similarity 56.9%; Pred. No. 4e-51;  
Matches 645; Conservative 0; Mismatches 485; Indels 4; Gaps 2;  
QY 1 ATGATTCAATTTTCAATTAATCGCACATTTATTTATTTATTCATGCTTTTAAATACAACTAAACGT 60  
Db 598 ATTAAGAATTAACCGTCTAAGCAAGTGTCTTTTACAGAAATTTACAACTGTTCAACA 657  
QY 61 GCTATTAGCACTAAATAATGCCATTCCTATTTCTTT-CATCAATAAATAATTTGAAGTCACTTC 119  
Db 658 GCAATTTCTTCTAAACAAACGATTCCTCAATCTTAAACAGGTGTTAAATTTGTGTATCAGA 717  
QY 120 TACAGGAGTAACTTTTACAGAGGTCTAAGGTCATATATCAATTTGAAACACTATTCCTGT 179  
Db 718 AGATGGCTTATCACTTACTGGGAGTAAACGCGATATTTCAATTTGAAGTATTTTAAAGTAA 777  
QY 180 AAGTAATGAAATGCTGTTTCTAATTAACCTCTCCAGAGCTATTTTATTTAGAACTAG 239  
Db 778 AGATGATGAAAGAGCTCAATGACCATTTGAACGACAGGTAGCATCGTTTACAACTCG 837  
QY 240 TTTTATTTAATATTTTCAAGTTTCCAGATATTTAGTATTAATTTTAAAGAAATTTGA 299  
Db 838 TTTCTTTGGCGAAATTTATTCGTAATTTACCAAGATATGTTTCACAATGGAAGTCTTACA 897  
QY 300 ACAACACAAAGTTGTTTAAACAGGTGTTAATCAGAGATTTACCTTTAAAGGAAAGATGT 359  
Db 898 TAACAATCAAGTTGCAATTTACTTTCAGGAAAGCTGATTTTACGGTTAATGGAATAGATGC 957  
QY 360 TGACAGATATCTCGTCTACAGAGATTTCAACAGAAATCTCTTTGATTTTAAAAACAAA 419  
Db 958 TGATTAATTTCCGCACTTACAGTAATTTGATCTCAAAACCAATTTGAATTTACCTGTTCA 1017  
QY 420 ATATTGAGTCTATTTATTTGCTGAAACAGCTTTTTCAGCCAGTTTTCAGAAAGTCTGCC 479  
Db 1018 TTTATTTGACAAAATTTATTTAGTGAACAGGTTTTCGTCTGATCGATCGACGAAAGTCTCC 1077  
QY 480 TATTTTACAGAGTTTCATATTTGATTAAGTAAATCATTAAGATTTTAAAGCAGTAGCGAC 539  
Db 1078 AATTTTAACTGGGTTTCACTTTATTTTGAATAATCAAAAA---TTACTTGGCGTTTCGAC 1134  
QY 540 TGACTCTCATCGTATGAGCCAAAGTTTAACTCACTTTGGACATATCTTCAGCAGATTGAT 599  
Db 1135 AGATTCATCGTTTAACTCAACGTTGATCCCGACAGAACAGCAGTAGAAGACCTTTAA 1194  
QY 600 GGTAGTTCTTCCAAAGTAAATCTTTGAGAGAAATTTTTCAGCAGTATTTTACAGATGATTTGA 659  
Db 1195 CATTGTAATTCAGGAAAAAGTTTAACTGAATTTTCTGTTTCTATTAAACCAATGAAGAAGA 1254  
QY 660 GACGTTGAGGTATTTTCTCACCAGCAAACTCTTTTTCAGAAAGTGAACACATTTCTTT 719  
Db 1255 AATGTTTGAATTCAGCATTTATGGAATTTCAAGTGTCTATTTTAAACAGAAACAAATGTACTT 1314  
QY 720 TTATACAGCTCTTTTAGAAGGAAATTTATCCCGATACAGACCGTTTATTAATCAGCAAAAT 779  
Db 1315 CTATTTCTGTTTGTAGAGGAAATTTATCTGATACCAACCGCTTTAATTTCCAATCTAGCCA 1374  
QY 780 TGAGACGAGGTGTTTCTTCAATACCCCAATCTCCGACGCTATGGAACGTCCTTCTTT 839  
Db 1375 TAAACGCAAAATTTGAATTTTATGTACAGAAATTTGCTTTTCAGCAATTCGAACGTCCTTT 1434  
QY 840 GATTCTTAATGCTACTCAAAATGGTACTGTTAAAGCTTGAGATTACTCAAAATCATATTTTC 899  
Db 1435 ACTTCTCATGAAGGACGTAAACATATTTGTTGCGCTTTTCAATTTTCCACCAATTTCTGTT 1494  
QY 900 AGCTCATGTTAACTCACCTGAGTTGTTAAGTAAACGAGGATTTAGATATTTGTTAGTCA 959  
Db 1495 TTTATATGAAATTTCACTGAAATTTGGAAGAGTTCGAAGAGCTTTTAACTATGAAATGT 1554

QY 960 GTCTGTAGTATTTAACTATCAGCTTCAATCAACTTACCTTATGAGTCTTTAAAGC 1019  
 Db 1555 TTCTGTGAAGCTTGGATATTTCTTCAACCCAGATATATGAAGATCGTTCGGAGC 1614  
 QY 1020 TATTAAAGTGAACAGTAAATTCATTTCTTATCAGCTTCAAGCTTCAACCTTAC 1079  
 Db 1615 GTTGGCGATATGAATATACCGTGAATTCCTTTCTCCAATTCGTCCTTTACATTTGA 1674  
 QY 1080 ACCAGCGATGAGGAAGTTTATCAATTAATTAACAGTGAAGCAAAA 1133  
 Db 1675 GCCACCGAAGCAAGTATTTCACTAATTAACCGGTTTCGTACAAA 1728

RESULT 34  
 LOCUS AR346291 1155 bp DNA linear PAT 17-AUG-2003  
 DEFINITION Sequence 902 from patent US 6583275.  
 ACCESSION AR346291  
 VERSION AR346291.1 GI:33743336  
 KEYWORDS Unknown.  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 1155)  
 AUTHORS Doucette-Stamm, L.A. and Bush, D.  
 TITLE Nucleic acid sequences and expression system relating to  
 Enterococcus faecium for diagnostics and therapeutics  
 JOURNAL Patent: US 6583275-A 902 24-JUN-2003;  
 FEATURES Location/Qualifiers  
 source 1..1155  
 /organism="unknown"  
 /mol\_type="genomic DNA"

ORIGIN  
 Query Match 28.0%; Score 317; DB 6; Length 1155;  
 Best Local Similarity 55.6%; Pred. No. 1.9e-48;  
 Matches 630; Conservative 0; Mismatches 500; Indels 3; Gaps 1;

QY 1 ATGATTCATTTTCAATTAATCGCACATTTATTTATTCATGCTTTTAAATACAACTAAACGT 60  
 Db 22 ATCATGAAGTTACTTTAAACCGAGTAGCTTTATCGAGGATTCGAACTGTTCAACGA 81  
 QY 61 GCTATTAGCACTAAATATGCTTCTTATTTTCAATCAATTAATTAATGAAGTCACTTCT 120  
 Db 82 GCTATTTCAAGCAAAACCAACGATCCCTATTTTGAACAGGTGTAATAATCACACTGACAA 141  
 QY 121 ACAGGAGTAACTTTAACAGGGTCTAACGCTCAATATCAATTTGAACACATTTCTCTGA 180  
 Db 142 GAAGTTTGACTTTGACGGGGAGCAACGCTGATATATCAATTTGAACATTTTGTCTGTT 201  
 QY 181 AGTAATGAATATGCTGTTGCTAATTAATCTCTCCAGGAGCTATTTTATTAAGAGCTAGT 240  
 Db 202 GAAACGAAAGCAAAATATGCAATCGAATCTACTGTTTCATTTGTTTACAGCAGT 261  
 QY 241 TTTTATTAATATTTATTTCAAGTTTGCAGATATAGTATATAATGTTAAAGAAATGAA 300  
 Db 262 TTTCTTTAGCGAATCAATTCGGAGACTTCTCGAAGAAACATTTACTTTAAGAGTTTGA 321  
 QY 301 CAACACCAAGTTGTTTAAACAGTGTAAATCAGAGATTTACTTTAAAGGAAAGATGTT 360  
 Db 322 AATAACCAAGTAGCGATCACTTCTGGAAGAGGAATTTATCGTAAATGGAATTAGATGA 381  
 QY 361 GACCAATATCTCTGCTCAAGAAGTATCAACAGAAAAATCTTTGATTTTAAACCAAAA 420  
 Db 382 GATAACTATCTCATCTTCTGTTGTCGAAGCCATAACAGATGAATTTACCTGTACAC 441  
 QY 421 TTTATGAAGTCTATTTATGCTGAACAGCTTTTGAGCCAGTTTCAAGAAGTCTGCT 480  
 Db 442 GTATTGACTAACTAATCAACGAAACAGTTTTTGTCTGCTCACAACATGAGAGTCTGCA 501  
 QY 481 ATTTTAAACAGGAGTTCATATTGATTAAGTAAATCAATAAGATTTTAAAGCAGTAGCGACT 540

Db 502 ATCTTGACAGGTGTCATTTCAATTATA---TCTGATAATCTTTTATTAGCTGTAGCTACT 558  
 QY 541 GACTCTCATCTATCAGCAAGCTTTATCTACTTTGGAACAATCTTTCAGCAGATTTGATG 600  
 Db 559 GATTCCTACCGCTTAAGTCAACGGGTGATCCAGTAGAACAAGCGGCTGATCATTTTGTAT 618  
 QY 601 GTAGTTCTTCCAAAGTAAATCTTTGAGAGAAATTTTCAGCAGATTTTACAGATGATATTGAG 660  
 Db 619 ATGTGTTATCTCGGAAAAGTTTGTATCGAATTAATCTCGCTCATTAACAAATGAAGAAGA 678  
 QY 661 ACCGTTGAGGTATTTTCTCAACAGCAAAATCTTTGTTTCAAGTGAACATTTCTTTT 720  
 Db 679 ATCGTCGAAATCAGCATTTATGAAAACCAAGTGTGTTTCAAAACAGAACGATGATTTTC 738  
 QY 721 TATACACGCTCTTAGAAGGAAATATCCCATACAGACCGTTTATTAATGACAGAAATTT 780  
 Db 739 TATTCCTGTTTGTAGAAGGAAATATCCAGATACCAATCGTTTGTATTTCTTCAAGTTT 798  
 QY 781 GAGACGGAGTTGTTTTCATACCAATCCCTTCGCCACGCTATGGAACGTCCTTCTTTG 840  
 Db 799 AATACAGAGTTGAATTTCTGTTCTAGCTTTTTCAGCGCGATCGAACGGCTTCTTTA 858  
 QY 841 ATTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTGAGATTAATCAAAATCATATTTCA 900  
 Db 859 CTCTTCATGAAGCCGCTTAACAACATCGTTCGTTTATCTATTCGTCAGATGCGGTTGTT 918  
 QY 901 GCTCATGTTAACTCACCTGAGGTTGGTAAGTAAACGAGGATTTAGATATTTAGTAGCAG 960  
 Db 919 TTATACGGAATTCGCGAAGTAAACGGAAGTGAAGAAAGCTTGAGTTATACAGCAAGC 978  
 QY 961 TCTGTAGTGAATTAACCTATCAGCTTCAATCCAACTTACCTTATTTAGTCTTTTAAAGCT 1020  
 Db 979 AGTGGCGATCCATTTAGATATTTCTTCAATCCGATTTATGAAAGCAGCGTTCGGTGA 1038  
 QY 1021 ATTTAAAGTGAACAGTAAATAATTCATTTCTATCAGCTTCAACAGTTTCACCTTACCA 1080  
 Db 1039 TTTGAGATATGAGCATCAAAAGTGAATTTATCTCTGCTATTTCTGCTTACATTTGAA 1098  
 QY 1081 CCAGCGCATGAGGAAGAAAGTTTATCCAAATTAATTAACAGTACGAAACAAA 1133  
 Db 1099 CCAAGGAAGTGGCGTTCAATTTATTCAGCTGATTTACTCTCTGACGTACAAA 1151

RESULT 35  
 LOCUS AL935252 343050 bp DNA linear BCT 21-FEB-2003  
 DEFINITION Lactobacillus plantarum strain WCFS1 complete genome; segment 1/11.  
 ACCESSION AL935252 AL935253  
 VERSION AL935252.1 GI:28269805  
 KEYWORDS Lactobacillus plantarum WCFS1  
 SOURCE Lactobacillus plantarum WCFS1  
 ORGANISM Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
 Lactobacillus.

REFERENCE 1  
 AUTHORS Kleerebezem, M., Boekhorst, J., van Kranenburg, R., Molenaar, D.,  
 Kuipers, O. P., Leer, R., Tarchini, R., Peters, S. A., Sandbrink, H. M.,  
 Fiers, M. W. E. J., Stiekema, W., Lankhorst, R. M. K., Bron, P. A.,  
 Hoffer, S. M., Groot, M. N. N., Kerkhoven, R., de Vries, M., Ursing, B., de  
 Vos, W. M., and Siezen, R. J.  
 Complete genome sequence of Lactobacillus plantarum WCFS1  
 Proc. Natl. Acad. Sci. U.S.A. 100 (4), 1990-1995 (2003)  
 MEDLINE 22480296  
 PUBMED 12566566  
 REFERENCE 2 (bases 1 to 343050)  
 AUTHORS Kleerebezem, M. and Siezen, R. J.  
 DIRECT SUBMISSION  
 TITLE Submitted (14-OCT-2002) Wageningen Centre for Food Sciences, P. O.  
 BOX 557, 6700 AN Wageningen, The Netherlands  
 JOURNAL Location/Qualifiers  
 FEATURES 1..343050  
 source 1. /organism="Lactobacillus plantarum WCFS1"  
 /mol\_type="genomic DNA"







```
gene      10196..10798
/gene="ssb"
/locus_tag="lp_0010"
CDS       10196..10798
/gene="ssb"
/locus_tag="lp_0010"
/codon_start=1
/transl_table=11
/product="single-strand binding protein"
/protein_id="CAD62710.1"
/db_xref="GI:28269813"
/db_xref="TrEMBL:Q890K1"
/translations="MINRTILVGRFLTRDPRLRYNGAAVAFTTIANRQFTNONGER
EADFIICVIRKAAEFANFTHKGSILVIGDRIOTRYENOCQVRVYTVVVFVPSL
LESRAEQRHQAANGSGNNYNNNGNSYNNNGYSGNQNAAPQSSANNNNPFGN
GNTGASSAAPPSSANNNNQDPFANNQIDISDDDLFF"
10837..11073
/gene="rpsr"
/locus_tag="lp_0011"
10837..11073
/gene="rpsr"
/locus_tag="lp_0011"
/codon_start=1
/transl_table=11
/product="ribosomal protein S18"
/protein_id="CAD62711.1"

Query Match      22.2%; Score 252; DB 1; Length 343050;
Best Local Similarity 53.3%; Pred. No. 5,1e-37;
Matches 606; Conservative 0; Mismatches 520; Indels 12; Gaps 3;

QY      2  TGATTCAATTTCAATTAATCGACATTAATTTATTCATGCTTTAAATCAACTAAACGTG 61
DB      1544  TTAATGAATTCAGATCAATCGATCTGCTTCATCAAGAAATTAATAACGTCCAACGG 1603
QY      62  CTATTAGCACTAAATGCAATTCCTATTCTTTCATCAATCAATTAAGTCACTTTCTA 121
DB      1604  CCAATTTCTCAAAACGACCATTCATTTTGACTGGCTTTAACTTGACGTTAATCTG 1663
QY      122  CAGGAGTAATTTAAACAGGCTCTAACGGTCAATATCAATTTGAAACACTATTTCTGTAA 181
DB      1664  AGCTATTACCTTAATCTGCTAGTGTGCTGATATTTCTATTGAAACCACTTCCAGCA 1723
QY      182  GTAATGAAATGCTGGTTGCTAATTAATCTCTCCAGGAGCTATTTTATTAGAGCTAGTT 241.
DB      1724  GTGATGACAAATACGTTAGTGTGAGATGCGCGTTCAATCGTTTACCTGCGCGTT 1783
QY      242  TTTTATTAAATATTATTTTCAAGTTTGCCAGATATTAGTATAAATGTTAAAGAAATTTGAC 301
DB      1784  TCTTTAGTGAATCGTAAATAATTTACAGAGACACCATGACTGTTTAACGTGTGCGAG 1843
QY      302  AACACCAAGTTGTTTAAACAGTGTGTAATACAGATTACCTTTAAAGGAAAGATGTTG 361
DB      1844  GTTTTCAACAAATTAATCTCAGGGCGGCTTCTTCAATCAACGGCTTAGACCGCTG 1903
QY      362  ACCAGTATCTCTGTCTACAGAAGTATCAACAGAAATTCCTTTGATTTTAAACAAAT 421
DB      1904  AAAACTACCCACACTTACTGAAATGATGACTAATACCATCACTAGCTAGCTGGGATG 1963
QY      422  TATTGAAGTCTATTATTCGTAACAGCTTTTGCCAGCGAGTTTACAGAAAGTCTGCTTA 481
DB      1964  TACTCAAAAGAGTTGATTTGGTCAGACCGTTATTGCGGTTTCCAATCAGAAAGTCGACCA 2023
QY      482  TTTTAAACAGGAGTTTCAATATGTTAGTATATCAAGATTTTAAAGCAGTACGAGCTG 541
DB      2024  TTCTGACCGGGTTTCAATTTATCTAGTAAACGGT---GAGTTCTTAGCGGTGCAACCG 2080
QY      542  ACTCTCATCTGATGAGCCCAAGCTTTTAAATCACTTTG---GACAACTCTCAGCAGATTGA 598
DB      2081  ATTACACCGGCTTAAGTCAACCGGCAATCAAGTTGCCAGAGCCAAATAGCCAAATTATG 2140
QY      599  TGGTAGTTCTTCCAGTAATCTTTGAGAGAAATTTTACGACGATTTTACAGATGATTTG 658
```

```
DB      2141  ATGTAATCATCTCTCGTAAAGTTTGACAGAAATTAATCTCGGATGATCGGACAAATAATC 2200
QY      659  AGACCGTTGAGGTATTTTCTCACCAGCCAAATCTTGTTCAGAAAGTCAACACATTTCTT 718
DB      2201  CTGACGTGCAGATCGGTTATCAGAAACCAAGTCTTATTTGTGCTGGGATACGTCCT 2260
QY      719  TTTATACAGGCTCTTAGAAGAAATTAATCCGATACAGACGTTTATTAATGACAGAAAT 778
DB      2261  TCTATTCTCGTCTACTCGAAGAAATTAATCCGACACGCTCAGCGTGAATCTCTAAAGAAT 2320
QY      779  TTGACACGAGGCTGTTTTCATATCCCAATCCCTTCGCGACCTATGGAAGTGCCTTCT 838
DB      2321  CTAATACAGCTTGTGAATTTTCAGCACCGGACCTTTGAGCAGCAATTTGACGGGCAATCC 2380
QY      839  TGATTTCTAATGCTACTCAAAATGGTACTGTT-----AAGCTTTCAGATTACTCAAAATC 892
DB      2381  TGTATATCCACGAAAGTCGCAATTAATGTGGTTTCCTCTCTCCGTTAACCCACGCAAAA 2440
QY      893  ATATTTACAGCTCATGTTAACTCAGCTGAGTGTGTAAGTTAAACGAGATTTAGATATTG 952
DB      2441  CGATTACGATTTTCGGTAACCTCACCAGATGTCGGGGAAGTAACGGAACAATACAGCCCA 2500
QY      953  TTAGTCAGTCTGTAGTGAATTAATCACTATCACTTCAATCCAACTTACCTTATTCAGTCTT 1012
DB      2501  CTGATTTGTCAGGTGACGAACCTTGAAATTTCAITTAACCCAGACTACATGAAGAAGCGT 2560
QY      1013  TAAAGCTATTAAAGTGAACAGTAAATAATTCATTTCTTATCACCAGTTTCGACCAATCA 1072
DB      2561  TACGTTCTGTTGGCAAGCTATGATCAAAATCTCATTCAGGATGCGCAATTCGGCCCAATCA 2620
QY      1073  CCCTAACACACGAGCGATGAGGAAGAAAGTTTATCCAAATTAATTAACACCAAGTACGAC 1130
DB      2621  CTTTAGTGCCAACTGAAGAAGGGAACCTTTATCAGTTGATTACCAAGCTTCGAAC 2678
```

```
RESULT 36
AB016744
LOCUS      300787 bp      DNA      linear      BCT 01-JAN-2003
DEFINITION      Staphylococcus epidermidis ATCC 12228, section 1 of 9 of the
complete genome.
ACCESSION      AB016744 AB015929
VERSION        AB016744.1 GI:27314460
KEYWORDS
SOURCE
ORGANISM      Staphylococcus epidermidis ATCC 12228
Staphylococcus epidermidis ATCC 12228
Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE      1 (bases 1 to 300787)
AUTHORS        Zhang, Y., Ren, S., Li, H., Fu, G., Lu, L., Lu, G., Jia, J., Tu, Y.,
Qin, Z., Chen, Z. and Wen, Y.
TITLE          Direct Submission
JOURNAL        Submitted (05-NOV-2002) Chinese National Human Genome Center at
Shanghai, 250 Bi Bo Road, Shanghai 201203, China
FEATURES
Location/Qualifiers
1..300787
/organism="Staphylococcus epidermidis ATCC 12228"
/mol_type="genomic DNA"
/strain="ATCC 12228"
/db_xref="ATCC:12228"
/db_xref="taxon:176280"
362..1717
/locus_tag="SE0001"
362..1717
/locus_tag="SE0001"
/codon_start=1
/transl_table=11
/product="chromosomal replication initiator protein"
/protein_id="AA003598.1"
/db_xref="GI:27314461"
/translational="MSEKIMDKVLEIAQERISNTSYQTFIKDTQLYSKNDEAILV
SLPPFASMLNQRYSIMQAIYDVIGYEVKPHFISEDELASVNNVTQVQPOVQHS
STDDKTWKEQFNMTFTFVIGPNRFPFAALVAEAPAEANPLFIYGVGLGK
THLMAIGHVLNKPNAKVIYTSSEKTFNEFKISIRDNETEAPREKRIKIDVLID
IQFIQNKQEQTEFFHTNELLQNNKQIVISSDRPPKEIAKLEDRSRFEWGLIVDI
```

```
gene
CDS
```



CDS  
15518..15847  
/locus\_tag="SE0010"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA03607.1"  
/db\_xref="GI:27314470"  
/translation="MTDLHTLVLIILCGVVLTLLRIVIPVMSIRVNLPAIVIKWLSF  
IPITFLTALIIDGVIQQHDHAFVLTNLPVILAIPTVPMVLAIFTRSLVTLIGGIFVI  
ACLRLLP"  
16136..17104  
/locus\_tag="SE0011"  
16136..17104  
/locus\_tag="SE0011"  
/codon\_start=1  
/transl\_table=11  
/product="putative homoserine-o-acetyltransferase"  
/protein\_id="AA03608.1"

Query Match 21.8%; Score 247.4; DB 1; Length 300787;  
Best Local Similarity 52.9%; Pred. No. 3.7e-36;  
Matches 580; Conservative 0; Mismatches 511; Indels 6; Gaps 2;

QY 1 ATGATTCATCAATTAATCGCATATTTATTCATCAATAAATAATTCAGTCACTTCT 60  
DB 2026 ATGATGGAATCAATTAAGAGATTTATTTAATCAACTTAACGACATTAAG 2085  
QY 61 GCTATTAGCACTAAATGCAATTCCTATTCATCAATAAATAATTCAGTCACTTCT 120  
DB 2086 GCTATCTCCTAGAACAACTTTACCGATTTTAACAGGTATCAAAATTTGATGCTAAGAA 2145  
QY 121 ACAGGATTAATTAACAGGTCTAAGGTCAAAATCAATTAATTAAGTCACTTCT 177  
DB 2146 ATGAAATCAATTTAAGTGGTTCAGATTCAGAAATATCAATAGAAATCACTTCCAAA 2205  
QY 178 GTAAGTAATGAAATGCTGTTGCTAATTTACCTCCAGGAGCTATTTTATAGAGCT 237  
DB 2206 CAAGTTGATGGTGAAGAAATTTGTAATTAAGAAACAGGATCAGTTGCTCTCGT 2265  
QY 238 AGTTTATTAATTAATTTCAAGTTTCCAGATATTTAGTATAAATGTTAAAGAAAT 297  
DB 2266 CGTTCTCGTTGATATTAATAAATCTCTCGAAAGAGTTAAATTAATTAATTAAT 2325  
QY 298 GAACAACACAGTCTTTTACCAGTGTGAATCAGAGATTAATTAATTAAGGAAAGAT 357  
DB 2326 GAACAATTCACAGTTAATCAATCAGTCAATCAGAAATTTAATTAAGTGGCTTAG 2385  
QY 358 GTTGACCAATCTCTGCTCAAGAGATATCAACAGAAATCTCTTTCATTTTAAACA 417  
DB 2386 CCGTATCAATACCAATTTATCTCAGGTATCAACAGATGACGCTATTAATTTGCTAG 2445  
QY 418 AAATTAATGAAGTCTATTTGCTGAACAGCTTTTGAGCCAGTTTACAAAGAAAGTCT 477  
DB 2446 AAGGTACTAAATAATATCAATGTCACAACTAATTTGCAAGTGTCCACCTCAGAAACGA 2505  
QY 478 CTTATTTAAACAGGAGTTCATATTCATTAAGTATCAATAAGATTTTAAAGCAGTACG 537  
DB 2506 CCAGTACTTACTGGGTAACTGGCTTATA--CAAGATATGAATTAATATGACAGCA 2562  
QY 538 ACTGACTCTCATCGTATGAGCAAGTGTATTAATCACTTTGGAATATCTTCAAGTATTTG 597  
DB 2563 ACAGATTCACACCGCTTAGCTGTAAAGATTACAGTTAGAAATGAATCAGAAATAA 2622  
QY 598 ATGGTAGTCTTCCAAAGTAATCTTTGAGAGAAATTTTCAAGCAGTATTTACAGATGAT 657  
DB 2623 AATGTCATCTCTGGTAAAGCTTTATCTGAATTAACAAATTAATGATGACAGCAC 2682  
QY 658 GAGACCGTTAGGATTTTCTCAACGAGCAATCTTTGTCAGAGTGAACATTTCT 717  
DB 2683 GAAGATATGATATTTCTTGTCTTCAACCAAGTGTATTCAGAGTGGGATATTAAT 2742  
QY 718 TTTTATACACCGCTTAGAGGAAATATCCGATACAGACCGTTTATTAATGACAGAA 777

Db 2743 TTCAATCTCAGGTTTACTTGAAGGTCATTTATCCAGATACGACAGGTTTATTCACAGAAAT 2802  
QY 778 TTTTGAGACGAGGTTGTTTCAATACCAATCCCTTCGCCACGCTATGGAAGTGCCTTC 837  
Db 2803 TATGAGATTAATTAAGGAATTAACATGAGACTTCTATCATGCAATTTGATCGTGCATCT 2862  
QY 838 TTGATTTCTAACTGCTACTCAAAATGCTACTGTTAAGCTTTGAGATTTACTCAAAATCATATT 897  
Db 2863 TTATTAGCAGCGTGAAGGTGGAATAATGTTTAAATTAAGTACAGGTAATGAATTAGTT 2922  
QY 898 TCAGCTCATGTTACTCACTGAGGTTGCTAAGGTAAGGATTAAGGATTTAGATTTAGTTAGT 957  
Db 2923 GAACTTTTCACTCTCTGAAATTTGCTGTTAAAGAAAGATTTAAACGCTAATGAT 2982  
QY 958 CAGTCTGCTAGTGAATTAACCTATCAGCTTCAATCAACTTACCTTATTTAGTCTTTTAAA 1017  
Db 2983 GTAGAAGCGGAAACTTGAATAATTTCTTCACTCAAAATCATGATGATGCTTTTAAA 3042  
QY 1018 GCTATTAAAGTGAACAGTAAATAATTCATTTCTTATCAACAGTTCGACCATTCACCTA 1077  
Db 3043 GCCATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3102  
QY 1078 ACACGAGCGGATGAGGA 1094  
Db 3103 AACCAAAAGATGATGA 3119

RESULT 37  
AE017200 300007 bp DNA linear BCT 26-FEB-2004  
LOCUS Lactobacillus johnsonii NCC 533, section 1 of 7 of the complete  
DEFINITION genome.  
ACCESSION AE017200 AE017198  
VERSION AE017200.1 GI:41582368  
KEYWORDS  
SOURCE Lactobacillus johnsonii NCC 533  
ORGANISM Lactobacillus johnsonii NCC 533  
Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
Lactobacillus.  
1 (bases 1 to 300007)  
Pridmore,R.D., Berger,B., Desiere,F., Vilanova,D., Barretto,C.,  
Pittet,A.C., Zwhalen,M.C., Rouvet,M., Altermann,E., Barrangou,R.,  
Mollet,B., Mercenier,A., Klaenhammer,T., Arigoni,F., and Schell,M.A.  
The genome sequence of the probiotic intestinal bacterium  
Lactobacillus johnsonii NCC 533  
Proc. Natl. Acad. Sci. U.S.A. 101 (8), 2512-2517 (2004)  
14983040  
2 (bases 1 to 300007)  
Pridmore,R.D., Berger,B., Desiere,F., Vilanova,D., Barretto,C.,  
Pittet,A.C., Zwhalen,M.C., Rouvet,M., Barrangou,R., Altermann,E.,  
Mollet,B., Mercenier,A., Klaenhammer,T., Arigoni,F. and Schell,M.A.  
Direct Submission  
Submitted (01-DEC-2003) Nutrition and Health, Nestle Research  
Center, P.O.Box 44, Lausanne 26, Vaud CH-1000, Switzerland  
FEATURES  
Location/Qualifiers  
1..300007  
/organism="Lactobacillus johnsonii NCC 533"  
/mol\_type="genomic DNA"  
/strain="NCC 533"  
/db\_xref="taxon:257314"  
1..1365  
/locus\_tag="LJ0001"  
1..1365  
/locus\_tag="LJ0001"  
/codon\_start=1  
/transl\_table=11  
/product="chromosomal replication initiator protein DnaA"  
/protein\_id="AA07981.1"  
/db\_xref="GI:41582369"  
/translation="MFDDJDFQFNAEMKKSYSTVAVNFWKTKDIPSNKTKEMI  
IAVSPVAKGYWEKNLASQLQEAIVADMEIQPFVAGKEGRLVTPQPRKTKNQ  
EILENRDEFADQLQNSKTFDFVQEGKNLAALAVADNPGSFYFNPLFPGGV  
GLGKTHLMQALGHQMLAEKPKAKVYVYIQSEFTFVNDINSIKNTKTOAEFRNRCDDL

LVDDIQFSSKKEGIEEFHPTETLYNDQKQIVMTSDRLPTEIPELSERLSRFAWGL  
QVEITPPDLETRIALRLKKAETDGLAIDDSTLDYIASQVDYNNIRELEGALVKQAHAT  
IERDINVDLAKEADLKLKVNKGLQSKIQEVVANYFOITSELKGRKRVKQIVV  
PRQIAMYSRELTDSSLPKIQEFGCKDHTTVMHACDKISRALKTDAEIKAAVYDLKA  
MLEH"

gene

1546. .2676

CDS

/locus\_tag="LJ0002"  
1546. .2676  
/locus\_tag="LJ0002"  
/codon\_start=1  
/transl\_table=11  
/product="DNA polymerase III, beta chain"  
/protein\_id="AAS07982.1"  
/db\_xref="GI:41582370"

/translation="WQFTINRMLFLENLNMRAISSRATIPILSGIKLTITDMLTLT  
TGSYDTSIEIQDNDYDLVOSTGIVLPARFFSEIVKLPFGKDFSEVKESFOTKI  
YSENTEFMINGLDANNYPHLPFEISTDASKISQKTFREIINETVPAVATQSRPRLTG  
VNFI PNNSISIKAVATDSHRLSORQISLENGPOTSLDII PGKSLVELSRIIGESPEI  
TVNPGENVLFEVGNIAFYSLLDQYPTDRLIPTESTTSVEPELVPALSLERASL  
LTHERNNVVMTLDVQVLKQDSPEIGNVEEIEGKNLEGDGLIISNPDPYLR  
ALRASITDSINMFTQPLRPFTVIPAKQDVNFTQITTPVTF"

gene

2951. .3100

CDS

/locus\_tag="LJ0002b"  
2951. .3100  
/locus\_tag="LJ0002b"  
/codon\_start=1  
/evidence=not experimental  
/product="hypothetical protein"  
/transl\_table=11  
/protein\_id="AAS07983.1"  
/db\_xref="GI:41582371"

/translation="MISSGQAKWYLVNPKVKNLGEVEDRRGKTIHPGDVLNLAGEEY  
EFISE"

gene

3104. .4228

CDS

/locus\_tag="LJ0003"  
3104. .4228  
/locus\_tag="LJ0003"  
/codon\_start=1  
/transl\_table=11  
/product="DNA replication and repair protein RecF"  
/protein\_id="AAS07984.1"  
/db\_xref="GI:41582372"

/translation="MYLANFELKDPNFEELKINPDHVNIFIGNAQKTNLEAIY  
FLALTRHRTSDKELIRFGSKAGQKVHKSQVEELKURLTPNGKKAWNRLQK  
KLSAYGVGNALFSPEDLALVKGAPSRRRRPMDFEQINSEYLYFLSQVQLVQOR  
NNYLKQLSIKKANDLFDVLSDQLAGIAAETISRRIKYIKLNSYAQSAHSEISGQA  
EKLOIFRYPKVEITPDDVETIYOKVITSYKKNRPNEIRKGTLSGPHRDOLDPLIN  
DKNADHFAOQORTISLSVLAETQLVHETQVEYPIILLDDVMSELDRHRQSSILNY  
IHGKTQITITTDLEGISWEIVKEPKVHISAGTISTKES"

gene

4229. .6196

CDS

/locus\_tag="LJ0004"  
4229. .6196  
/locus\_tag="LJ0004"  
/codon\_start=1  
/transl\_table=11  
/product="DNA gyrase subunit B"  
/protein\_id="AAS07985.1"  
/db\_xref="GI:41582373"

/translation="MAEDKSKAALKAKDFEKRAPTYNASOIOVLGGLLEAVRKRP  
IGSTSSQGLHVLWETIIONSIDERLAGFATKLEITVNEDEGVTYVDDGCGIPVDLOAK  
TGRPALEVTFTVLHAGGFGGGYKVGSLHGVGASVNALSTKLEVTVMRDGKKYAI  
SPDGRVRVGMKQVTVPLBEHGTITVRFYPPDIFTEFTTDDKILKNRIELAPLNK  
GLKLTFTDKRESAETVYHEGGIKYEVAFNLKQOEVLFDPEIYVEGNYEGIDVEVA  
LOYTNGYKTIITLMFANNTHTEVGGHEAGFKTALTRVVDYAHKAKILKDDNLNLSGI  
DIREGITAIVSKHPQPEGOTKTKLGNSDARTAVDRAFSETFSRFLMENPNSVGRKI  
VEKGQLAERARTAAKAREVTRKKSGLAIANLPGLADNTSNDPSISELFIVEGNSAG  
GSAKGRSLTQAILPIRGKILNVKSAKMDRILANKEYSUTALGTGFGADFDVSKA  
RYHKLITMDADVDAHIRTLLLTFLFYNMRPMEIKGYVYIARPLYQVROCKVEYL  
TDDEBLHGLTQPSKXPIVQRYKGLGEMDPQEQWETTMNPNERRLRDRVSPYAEADA  
DKIFOLLGMNEVGPREFTEKNAKYVENLDA"

gene

6207. .8696

CDS

/locus\_tag="LJ0005"  
6207. .8696

/locus\_tag="LJ0005"  
/codon\_start=1  
/transl\_table=11  
/product="DNA gyrase subunit A"  
/protein\_id="AAS07986.1"  
/db\_xref="GI:41582374"  
/translation="MANENQPDHRIIRNVDLTHTMRTSFLDYAMSVIVARALPDVRDG  
LKPQRRILYGMHGLGVTDPQKYKKSARIIVGEVMGKFPHGDSYIYLAHAHWAQDFSY  
RYMLVDGHGNGFSGDGPAAWRYTEAKMSIAVEMLRDINKETIDWQRNYDDTENEP  
VVLPAFIPNLNVGASGIVAGVMTTNI PPHNLAIEVIRGHMLMDNPVTTKOLMKVIPG  
DPFTGTGIVMGREGGIIRAYETKGNIVVRAKTEIETEKNGRERIVVTELPYLVNKAEL  
VKKIADLAREKTIIDIGTVRDESQDGMRTITIDIRRDASAVLVNLFKETOQANFG  
MMVAIVNGAPHLTKOMLEYLHODEVITRPTKFKELKKAERAAHILAGLRIALDH  
IDEIKIRGSONSDIAQILQNFGLDDRSQAILDMRLVRLTGLERDKIAEABQDL  
QAKIADYKDIILAKPERIDEIILYNELLDIQKRRDDRRRTKIAEGEAVSEDEDLIEOKD  
LLVLTTHNGYETIKRMPAEFKVQNGRGKIKGQVODDDFINHLIFSSDHYDLFFVTNL  
GKVTSKAYETPEFRNAKGLPIVNLLELDKGEKIQAIVNPEGADNHLFFVTMGT  
VRKTVSEFQNIIRRSGLIALTRDQDELNNVLTGDKNIGLITGHLYAVTFENEKQVR  
SMGRTAAGVGINRDNDYVVGSDILEPDSVLVISEKGYGKRTAASEYPVKGRGKG  
IKTANITEKNGPLAGVTVVGDEDEIMLITSAGVMIRFDVDDVSGTGRATLGVRLIKVD  
DGAQVASITAVPKASDEDEESTABEEPVEN"

gene

8883. .9179

CDS

/locus\_tag="LJ0006"  
8883. .9179  
/locus\_tag="LJ0006"  
/codon\_start=1  
/transl\_table=11  
/product="30S ribosomal protein S7"  
/protein\_id="AAS07987.1"  
/db\_xref="GI:41582375"  
/translation="MAKTYEVTYIKPDIDEDKALIDRFDKVVTDNGAELESKD  
WGRKFAYEIKYREGTYIHMTFVAENSEPVEFGRLSRIDNQILRSMTVKLDK"

gene

9213. .9731

CDS

/locus\_tag="LJ0007"  
9213. .9731  
/locus\_tag="LJ0007"  
/codon\_start=1  
/transl\_table=11  
/product="single-strand binding protein"  
/protein\_id="AAS07988.1"  
/db\_xref="GI:41582376"

/translation="MINNVLVGLTRDPDLRTTSGGISVATFTLAVDROYTNSOGER  
GADTISCIVIRKAAENFANFTSKGSLGICIGRIOTRTYDDKDGKEAVYTEVLVDNFSL  
LESARDRENROTNGENFAPQGNAPSTNFGSSAPSMNAPAGESKNKQDPFADSG  
STIISDDDLFP"

gene

9756. .9992

CDS

/locus\_tag="LJ0008"  
9756. .9992  
/locus\_tag="LJ0008"  
/codon\_start=1  
/transl\_table=11  
/product="30S ribosomal protein S19"  
/protein\_id="AAS07989.1"  
/db\_xref="GI:41582377"

/translation="MAQQRGRRRRKVDFAANHIDYDKVDVLLKRFISERGKIL  
PRRVGTSAKQKRLTVAIKRVRVGLLFFVAED"

gene

10086. .11003

CDS

/locus\_tag="LJ0009"  
10086. .11003  
/locus\_tag="LJ0009"  
/note="Contains Mg+ and Co2+ transporter domain"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAS07990.1"  
/db\_xref="GI:41582378"

/translation="MIKQKLAGNIDFYKMYDIYVNCDDHDIRLKKDQDLTSEIISY  
ITDHERPHDHYDSDLLVYDVPVMTADADFTLPIKFLMVDHTLFTVHSPDT  
TYMDEEQRDEHIHSEKELIFAILFAVTKYFORALSQLNSORLVLDNNLSERHKN  
DQELSOVEKSLVLSRSSIRTNLMLESKKNKSGLHWNASEECCDDIIIEVQOSSQ  
MKIYSEVTEISIKTSLNNLNNTMQFLTIVWSLLLTIFTIVTGGFGGMVDLPPIH  
NFFDWVIIIVGSIIVMAILWYLRHNNML"

```
gene      11110..13128
CDS       /locus_tag="LJ0010"
          11110..13128
          /locus_tag="LJ0010"
          /note="contains DHH phosphoesterase domain"
          /codon_start=1
          /evidence=not_experimental
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="AAS07991.1"

Query Match      21.3%; Score 242; DB 1; Length 300007;
Best Local Similarity 53.4%; Pred. No. 3.6e-35;
Matches 609; Conservative 0; Mismatches 510; Indels 21; Gaps 4;

QY 1 ATGATTCAAATTTTCAATTAATCGCACATATTTATTCATGCTTTAAATCAACTCAACAACTAACCGT 60
DB 1543 ACGATGCGATTTTCAATTAATCGTAATTTATTCCTCGAATACCTTAATATGCTATGCGT 1602

QY 61 GCTATTAGCACTAAATGCGATTCCTATTCCTTCATCAATAAATAATGAAGTCACTTCT 120
DB 1603 GCAATTTCTTCAGTGTCTACTATTCCATTTAAGTGTATTAATACTTACCCTTACAGAT 1662

QY 121 ACAGGAGTAATTTAAACAGGTCTAACGGTCAATATCAATTCATAAACAATTCCTTCTGTA 180
DB 1663 GAAATGCTTACTTTAACCGTAGTGATCTGACATTTCTATTGAAATTCAAATCCAGT- 1721

QY 181 AGTAATGAAATGCTGGTTGCTAAATTAACCTCTCCAGGAGCTATTTTATTAGAAGTACT 240
DB 1722 -----AAACGATGATTTAATTTGTTCAATCTACGGGATCGATCGTTTACAGACGA 1773

QY 241 TTTTATTAAATATTTTCAAGTTTGCAGATATTTAGTATTAATATTTAAAGAAATGAA 300
DB 1774 TTTTTCAGTAAATTTGTAATAAATTTACCAGGTAAAGATTTTTCATTTGAAGTAAAGAA 1833

QY 301 CAACACCAAGTGTGTTTAAACCAAGTGTGTAATCAGATTAACCTTAAAGGAAAGATGTT 360
DB 1834 AGTTTCAAACGAAATTTGTTCTGAAATACCGAATTCATGATTAACGTTTAGATGA 1893

QY 361 GACCACTGCTGCTCAACAGAGTATCAACAGAAATTCCTTTGATTTTAAACAAATA 420
DB 1894 ACAATTTATCCACATTTACAGAAATTTCTACTGATGATCATTTCAAATTTCTGTAA 1953

QY 421 TTATTGAAGTCTATTTATCTGCAACAGCTTTTGCAGCAGTTTACAGAAAGTGTCT 480
DB 1954 ACATTTAGAGAAATTTATTAACGAAATGTTTTTGCAGTGTCTTCAAGAAAGTGTCTCA 2013

QY 481 ATTTTACAGGAGTTCATATTTAAGTATCAATAAGATTTTAAAGCAGTAGCGACT 540
DB 2014 ACTTTAACTGGTTAACTTTATCTT---CAACAATTCATCAATTAAGCAGTTGCTACC 2070

QY 541 GACTCTCATCGTATGAGCCAGTTTAACTTAACTTCTGGAATTA---CTTCAGCAGATTG 597
DB 2071 GACAGTCACCGACTATCTCAAGCTCAAAATTTCTCTAGAAATATGGTCTCTCAACAACT 2130

QY 598 ATGTAGTGTCTTCCAAAGTAAATCTTTGAGAGAAATTTTCAGAGATTTTACAGATGAT 657
DB 2131 GACTTAAATTTCTCGGAGAGAGCTTAGTAGAATATCCGAAATATCGGAGAAAGTAT 2190

QY 658 GAGACCGTGTAGGTATTTTCTCACCAGCCAAATCTTTGTTTCAAGTGAACACATTTCT 717
DB 2191 CCTGAAATTCACAGTAAATCTGTTGTAACCAAGTTTATTTGAAGTTGGAACATTTGCA 2250

QY 718 TTTTATACAGCGCTCTTAGAGGAAATTTATCCGATACAGACCTTTTATTAAAGACAGAA 777
DB 2251 TTCATTACCGTTTACTTGTAGGACATATCCAGACACTGATCGTTTAAATTTCCAACTGAA 2310

QY 778 TTTGAGACGAGGTGTTTTCATTAACCAATFCCCTTCGCAACGCTATGGAACGTCCTTC 837
DB 2311 TCTCAACATCAGTTGAATTTGAATTTACCAAGTTTGTAGTCTGTTCTCTCGAAGTGCAGT 2370

QY 838 TTGATTTCTAATGCTACTCAAAATGGTACTGTTAAG-----CTTGAGATTACTCAAAAT 891
```

```
DB 2371 CTTCTTACTCACGAAGCGGTAATAATGTGGTTAAGATGACTCTTGTATGTTCAAATCAA 2430
QY 892 CATATTTTCAGCTCATGTTAACTCACTCGAGGTGGTAAAGTAAACGAGGATTTAGATATT 951
DB 2431 TTAGTTAAACTCCAAGGTGATTCACCGGAATAGGAATGGGAGAGAANAATTTGGCTTT 2490
QY 952 GTTACTAGTCTGGTAGTGATTTAACTATCAGCTTCAATCAACCTTACCTTTATTGAGTCT 1011
DB 2491 AAGAATCTTGAAGGAGATGGCTTAACTATTTCTTCAATCTGATTTATCTAAGAGAAGCT 2550
QY 1012 TTAAAGCTATTAAAGTGAAACAGTAAATAATTCATTTCTTATCACCAAGTTCCGACCATTC 1071
DB 2551 TTACGTGCTCTATTACTGATTTCCATTTATTAAGACTTTACCCAGCGCTAAGACCATTT 2610
QY 1072 ACCCTAACACAGGCGATGAGGAAGAAAGTTTTCATCAATTAATTAACCAAGTACGAACA 1131
DB 2611 ACAGTATACCGGCAAAACAAGATGTTAATTTCACTCAATTAATCACGCCAGTTAGAAC 2670

RESULT 38
AX926711
LOCUS      349980 bp      DNA      linear      PAT 19-DEC-2003
DEFINITION      Sequence 1 from Patent WO03084989.
ACCESSION      AX926711
VERSION      AX926711.1 GI:40246927
KEYWORDS
SOURCE      Lactobacillus johnsonii
ORGANISM      Lactobacillus johnsonii
Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
Lactobacillus.
REFERENCE
1
AUTHORS      Pridmore, R.D., Mollet, B., Arigoni, F. and Hermanns, J.
TITLE      Lal - the genome of a lactobacillus strain
JOURNAL      Patent: WO 03084989-A 1 16-OCT-2003;
SOCIETE DES PRODUITS NESTLE S.A. (CH)
FEATURES
source
1. 349980
/organism="Lactobacillus johnsonii"
/mol_type="unassigned DNA"
/db_xref="taxon:33959"
/notes="original seq 1 to long, 1983043 bases-divided as 7
new seq-seq 1 from 0.000.001 to 0.349.980-seq 2 from
0.300.001 to 0.649.980-seq 3 from 0.600.001 to
0.949.980-seq 4 from 0.900.001 to 1.249.980-seq 5 from
1.200.001 to 1.549.980-seq 6 from 1.500.001 to
1.849.980-seq 7 from 1.800.001 to 1.983.043"

ORIGIN
Query Match      21.3%; Score 242; DB 6; Length 349980;
Best Local Similarity 53.4%; Pred. No. 3.5e-35;
Matches 609; Conservative 0; Mismatches 510; Indels 21; Gaps 4;

QY 1 ATGATTCAAATTTTCAATTAATCGCACATTTATTTATTCATGCTTTAAATAACAATAAAGCT 60
DB 1543 ACGATGCGTGTACAAATTAATCGTAATTTATTTCTCGAATAACCTAAATAATGCTATGCT 1602

QY 61 GCTATTAGCACTAAATAATGCCATTCCTATTTCTTTCATCAATAAATAATGAAGTCACTTCT 120
DB 1603 GCAATTTCTTCAGTGTCTACTATTCCATTTAAGTGTATTAATACTTACCCTTACAGAT 1662

QY 121 ACAGGAGTAACTTTAAACAGGTCTAACCGTCAATATCAATTTGAAACACATTTCTCTGTA 180
DB 1663 GAAATGCTTACTTTAAACCGTAGTGATCTGACATTTCTATTGAAATTCAAATCCAGT- 1721

QY 181 AGTAATGAAATGCTGGTTTGTCTAATTAACCTCTCCAGGAGCTATTTTATTAGAAGTACT 240
DB 1722 -----AAACGATGATTTAATTTGTTCAATCTACGGGATCGATGTTTACCGACGA 1773

QY 241 TTTTATTAAATATTTTCAAGTTTGCAGATATTTAGTATTAATATTTAAAGAAATTTGAA 300
DB 1774 TTTTTCAGTAAATTTGTAATAAATTTACCAGGTAAAGATTTTTCATTTGAAGTAAAGAA 1833

QY 301 CAACACCAAGTGTGTTTAAACCAAGTGTGTAATCAGATTAACCTTAAAGGAAAGATGTT 360
```

```

Db 1834 AGTTTCAAACGAAATTTCTTCTGAAATATACCGAATTCATGATTAACGTTTAGATGCA 1893
Qy 361 GACCAAGTATCTCGCTCAACAAGAGTATCAACAGAAATCCCTTTGATTTTAAACAAAAA 420
Db 1894 AACAAATATCCACACTTACCAGAAATTTCTACTGATGATCATTCATAAATTTCTGGTAA 1953
Qy 421 TTATTTGAAGTCTATTATTTGCTGAACAGCTTTTGGAGCAGTTTCAAGAAAGTCTGCTCT 480
Db 1954 ACATTTAGAGAAATTAATTAACGAAATCTGTTTTGCGAGTCTACTCAAGAAAGTCTGCA 2013
Qy 481 ATTTTAAACAGGAGTTCATATTTCTTATTAAGTATCATATAAGATTTTAAAGCAGTACGACT 540
Db 2014 ACTTTAACTGGGTAACTTTATCTTTATCTT---CAACAATTCATCAATTAAGCAGTCTGCTACC 2070
Qy 541 GACTCTCATCGTATGAGCAACGTTTAAATCACTTTTGGACAATA---CTTCAGCAGATTTG 597
Db 2071 GACAGTCACCGACTATCTCAACGTCATAATTTCTCTAGAAATGGTCTCTCAACAGTACT 2130
Qy 598 ATGGTAGTCTTCCAGTAATATCTTTGAGAGAAATTTTTCAGCAGTATTTACAGATGATATT 657
Db 2131 GACTTAAATTAATCTCGGAAGAGCTTAGTAGAATTAATCCGAAATTAATCGGAGAAAGTGAT 2190
Qy 658 GAGACGCTGAGGTATTTTCTCACCAGCCAAATCTTGTTCAGAGTGAACACATTTCT 717
Db 2191 CCGTGAATCACAGTAATCTCGTGGAACCAAGTTTATTTGAAAGTTGGAAACATGCA 2250
Qy 718 TTTTATACAGCGCTCTTAGAGGAAATTAATCCCGATACAGACCGTTTATTAATGACAGAA 777
Db 2251 TTCTATTCAGGTTACTTGTAGGACAAATATCCAGACATGATCGTTTAAATCCAACTGAA 2310
Qy 778 TTTGAGACGAGGTTGTTTTCATATCCCAATCCCTTGGCCAGCTATGGAAGCGCTTC 837
Db 2311 TCTCAACATCAGTGAATTTGAAATTAACGATTTTAGCTGCTCTCTCGAAGCTGCCAGT 2370
Qy 838 TTGATTTCTAAGTCTACTCAAAATGCTAGTTAAG-----CTTGAGATTAATCAAAAT 891
Db 2371 CTCTTACTCAGAAAGCGTAATTAATGGTTTAAAGATGAGTCTTGATGTTCAAAATCAA 2430
Qy 892 CATATTTTCAAGTCTATGTTAACTCAGCTGAGGTTGGTAAAGTAAAGAGGATTTAGATATT 951
Db 2431 TTAGTTAAATCTCAAGTGATTTCCCGGAAATAGGAAATGGGAAGAGAAATGGCTTT 2490
Qy 952 GTTAGTCTGAGTGTAGTATTAATCAATCAGCTTCAATCCAACTTCACTTATGAGTCT 1011
Db 2491 AAGAATCTTGAAGGAGATGGCTTAACTATTTCTTCAATCTCTGATTTATCTAAGAGAAGT 2550
Qy 1012 TTTAAAGCTATTAAAGTGAACAGTAAATTAATTTCTTATCACCAGTTCGACCATTC 1071
Db 2551 TTACGTGCTCTATTACTGATTTCCATTTATTAAGAACTTTTACCAGCGCTAAGACCATTT 2610
Qy 1072 ACCCTAAACACCGGCGATGAGGAAGAAAGTTTATCCCAATTAATTAACACAGTACGAA 1131
Db 2611 ACAGTATACCGGCAACAAAGATGTTAATTTCACTCAATTAATCACGCCAGTTAGACA 2670

```

```

RESULT 39
LOCUS AX6211126
DEFINITION Sequence 4089 from Patent WO02094868.
ACCESSION AX6211126
VERSION AX6211126.1 GI:28449882
KEYWORDS
SOURCE Staphylococcus aureus
ORGANISM Staphylococcus aureus
Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE 1
AUTHORS Massignani,V.C., Mora,M.C. and Scarselli,M.C.
TITLE Staphylococcus aureus proteins and nucleic acids
JOURNAL Patent: WO 02094868-A 4089 28-NOV-2002;
Chiron Spa (IT)
FEATURES
source Location/Qualifiers
1..1131

```

```

/organism="Staphylococcus aureus"
/mol_type="unassigned DNA"
/db_xref="taxon:1280"

```

ORIGIN

```

Query Match 20.8%; Score 236.2; DB 6; Length 1131;
Best Local Similarity 52.2%; Pred.No. 1.3e-33;
Matches 573; Conservative 0; Mismatches 518; Indels 6; Gaps 2;
Qy 1 ATGATTCATATTTCAATTAATCGCACATTTATTTATTCATGCTTTTAAATACAACTAAACGT 60
Db 1 ATGATGGATTTCACTATTTAAAGAGATTTATTTTACACAATTAATGACACATTAAAA 60
Qy 61 GCTATTAGCACTAAAAATGCCATTCTTATTTTTCATCAATAAAAAATGAAAGTCACTTCT 120
Db 61 GCTATTTCCCAAGAAACAACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 120
Qy 121 ACAGAGTAACCTTTAAACAGGCTTAACGGTCAATATCAATTAATTAATTAATTAATTAATTA 180
Db 121 CATGAAGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 180
Qy 181 AGT---AATGAAATGCTGGTTTGTCTAATTAACCTCCAGGAGCTATTTTATTAAGAGCT 237
Db 181 ACTGTAGATGCGAGATATTGTCAATTTTCAGAAACAGGCTCAGTAGTACTTCTCTGA 240
Qy 238 AGTTTTTTTATTAATTAATTTCAAGTTTGGCAGATATTAGTATTAATTTTAAAGAAATT 297
Db 241 CGATTTCTTTGTTGATTTATTAATAAATAATTAATTAATTAATTAATTAATTAATTAATTA 300
Qy 298 GAACAAACCAAGTTGTTTAAACAGTGGTAAATCAGAGATTAACCTTAAAGGAAAGAT 357
Db 301 GAACAAATCCAGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 360
Qy 358 GTGACAGTATCTCGTCTACAAGAGTATCAACAGAAATTCCTTTGATTTTAAACAA 417
Db 361 CCAGATCAATATCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 420
Qy 418 AAATTAATTAAGTCTATTAATTTGCTGAACAGCTTTTGGCCAGTTTACAGAAAGTCTGT 477
Db 421 AAAGTCTTAAACAAAGCTGATTTGCAACAAATTTTGCAGTGTCCACCTCGAAGAACAGC 480
Qy 478 CCTATTTAAACAGGATTCATATTGTTAAGTATCAATAAGATTTTAAAGCAGTACGG 537
Db 481 COAGTACTAATCTGGTGTGAATCTGGCTTATA---CAAGAAATTAATTAATTAATTAATTA 537
Qy 538 ACTGACTCTCATCGTATGAGCCAAAGTTTAACTTTTGGCAATATCTTTCAGCAGATTTG 597
Db 538 ACTGACTCACACCGCTTGGCTGTAAAGATTTGAGTATTAAGATTTTCTGAAACAAA 597
Qy 598 ATGGTAGTCTTCCAAAGTAAATCTTTTGAAGAAATTTTTCAGCAGTATTTACAGATGATTT 657
Db 598 AATGTCTATCAATTCAGGTAAAGCTTTAGCTGAATTAATTAATTAATTAATTAATTAATTA 657
Qy 658 GAGACGCTGAGGTATTTTCTCAGCAAGCCAAATCTTTGTTTCAAGAGTGAACACATTTCT 717
Db 658 GAAGACATTTGATATCTCTTTGCTTCAACCAAGTTTATTTTAAAGTTGGAAATGTGAAC 717
Qy 718 TTTTATACACGCTCTTAGAAGGAAATTAATCCGATACAGACCGTTTATTAATGACAGAA 777
Db 718 TTTATTTCTCGATTTTAGAAGGACATTTATCTGATACACACGTTTATTTCCCTGAAAC 777
Qy 778 TTTGAGACGAGGTTGTTTTCAATACCAATCCCTTCCCGCAGCTATTCGAAAGTGCCTTC 837
Db 778 TATGAAATTTAAATTAAGTATAGACAAATGGGAGTTTTATCATGCGATTTGATCGTCTCT 837
Qy 838 TTGATTTCTAATGCTACTCAAAATGGTACTGTTAGCTTTGAGATTTTACAAATCATATT 897
Db 838 TTATTTAGCGCGTGAAGGTTGTTAAACGTTTATTAATTAATTAATTAATTAATTAATTAAT 897
Qy 898 TCAGTCTATGTTAACTCACTGAGGTGGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 957
Db 898 GAAATGCTTCTTACATCAACGAAATTTGTTACTGTAAAGAAAGTAAAGTAAAGTAAAGTAA 957

```



QY	958	CAGTCTGGTAGTGAATTTAACTATCAGCTTCAATCCAACTTACCTTATTGAGTCTTTAAAAA	1017
Db	958	GTTGAAGGTTGGTAGCCTGAAATTTTCATTCAACTCTAAATATATGATGGATGCTTTAAAA	1017
QY	1018	GCTATTAAAGTGAACAGTAAATAATTCATTCTTATCACAGTTCGACCAATTCACCCCTA	1077
Db	1018	GCAATCGATATGATGAGGTGAAGTTGAATTTCTCGGTACAATGAACCAATTTATTCTA	1077
QY	1078	ACACGAGCGGATGAGGA	1094
Db	1078	AAACCAAAGGTGACGA	1094
RESULT 40			
AX554149			
LOCUS	AX554149	1134 bp	DNA linear PAT 27-NOV-2002
DEFINITION	Sequence 1 from Patent WO0244718.		
ACCESSION	AX554149		
VERSION	AX554149.1	GI:25898086	
KEYWORDS			
SOURCE	Staphylococcus aureus		
ORGANISM	Staphylococcus aureus		
REFERENCE	Bacteria; Firmicutes; Bacillales; Staphylococcus.		
AUTHORS	Pelletier, J., Gros, P. and Dubow, M.		
TITLE	Compositions and methods involving an essential Staphylococcus aureus gene and its encoded protein staaU-r2		
JOURNAL	Patent: WO 0244718-A 1 06-JUN-2002;		
	Phagotech Inc. (CA)		
FEATURES	Location/Qualifiers		
source	1..1134		
	/organism="Staphylococcus aureus"		
	/mol_type="unassigned DNA"		
CDS	/db_xref="taxon:1280"		
	1..1134		
	/note="unnamed protein product"		
	/codon_start=1		
	/transl_table=1		
	/protein_id="CAD58216.1"		
	/db_xref="GI:2589808"		
	/translation="MMEFTRKDYFTQLNDTLKASPRITLPILTIKIDAKHEVI FTGSDSEHSIEITPKTVGDGIVNISETGSVLPGRFPVDIIKLPKGVKLSNEQ FOTLITSGSEFNLSLOPDQYVLLPOVSRDDAIQLSVKLVKVAQTNFVASTSETR PVLTVGNHIIQENELICTATDSHRLAVKQLQEDVSEKNVLIIPGKLAELNFKIMSD EDDIIPFASNOVLKVGNNPISRLLEGHYDPTLRFPENYEIKLSDNGEYHAI RASLLAGGNNVILKLGDDVVELSSTSPEIGTVKEVDANDVEGSLKTSFNSKYM MDALKAIIDNDEVEVEFFGTGMPKPFILKPKGDDSVTQLILPRTY"		
ORIGIN			
Query Match	20.8%;	Score 236.2;	DB 6; Length 1134;
Best Local Similarity	52.2%;	Pred. No. 1.3e-33;	
Matches 573;	Conservative 0;	Mismatches 518;	Indels 6; Gaps 2;
QY	1	ATGATTCAATTTTCAATTAATCGCACATTTATTATTTCATGCTTTTAAATACAACTAAACGT	60
Db	1	ATGATGGAAATTCACATTTAAAGAGATTTATTTTATACAAATTAATGACACATTAAAA	60
QY	61	GCTATTAGCACTAAAAATGCCATTTCCTATCTTTTCATCAATAAAAAATTTGAAGTCACTTCT	120
Db	61	GCTATTTTACCACGAAGCAACACATTACCTATATTAACCTGATCAAAATCGATGCGAAGAA	120
QY	121	ACAGCAGGTAACTTTAAACAGGGTCTAACGGTCAAAATCAATCAATGAAAACTATTTCCTGTA	180
Db	121	CATGAAGTTATTTAACTCTGGTTCCAGCTCTGAAATTTCAATGAGAAATCACATTTCCTAAA	180
QY	181	AGT----AATGAAAAATGCCTGGTTTGTCTTAATTAACCTCTCCAGGAGCTATTTTATTAGAAGCT	237
Db	181	ACTGTAGATGCGGAAGATTTGTGTCATATTTTCAGAAAAACAGGCTCAGTAGTACTTCTCTGA	240
QY	238	AGTTTATTTTATTAATTTATTTTCAAGTTTGCAGATATTAGTATATAAATGTTTAAAGAAAT	297
Db	241	CGATCTCTTGTGATTTATTAATAAAAAATTTACTCGTGAAGATGTTTAAATTTATCTACAAAT	300

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2005, 10:52:07 ; Search time 2733 Seconds  
(without alignments)  
15119.889 Million cell updates/sec

Title: US-10-048-071-27  
Perfect score: 1134  
Sequence: 1 atgattcaatttcaattaa.....ttacaccagtcagcaaac 1134

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST: \*  
1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_hic: \*  
4: gb\_est3: \*  
5: gb\_est4: \*  
6: gb\_est5: \*  
7: gb\_est6: \*  
8: gb\_gsei: \*  
9: gb\_gsei2: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	66	5.8	1101	CNS0039G	AL063921 Drosophila
2	63.6	5.6	603	BZ296086	BZ296086 CG1849.f1
3	63.6	5.6	614	BZ295031	BZ295031 CG1245.f1
4	57.8	5.1	1101	CNS008VL	AL069706 Drosophila
5	57.2	5.0	1043	CNS0145P	AL103735 Drosophila
6	57	5.0	429	CG803197	CG803197 1118039H0
7	56.2	5.0	1101	CNS008VL	AL069706 Drosophila
8	56.2	5.0	1225	CNS0161D	AL106171 Drosophila
9	55.6	4.9	1101	CNS008VL	AL071206 Drosophila
10	55.4	4.9	738	BX159596	BX159596 Danio rer
11	55.4	4.9	1036	CD387684	CD387684 AGENCOURT
12	55.2	4.9	876	CNS009G1	AL053529 Drosophila
13	55	4.9	1101	CNS0039G	AL063921 Drosophila
14	54.8	4.8	1200	CNS016CO	AL106578 Drosophila
15	54.6	4.8	705	BH941456	BH941456 odg40c07
16	54.4	4.8	1061	CNS015LM	AL105604 Drosophila
17	54	4.8	987	CNS014PQ	AL104456 Drosophila
18	54	4.8	1092	AL175696	AL175696 Tetraodon
19	53.8	4.7	471	CG807194	CG807194 1118078H0
20	53.6	4.7	1101	CNS008BFA	AL069119 Drosophila
21	53.6	4.7	1101	CNS016LI	AL106896 Drosophila
22	53.4	4.7	1092	CNS020K7	AL175696 Tetraodon
23	53	4.7	969	CL469446	CL469446 SAIL_1303
24	52.8	4.7	1101	CNS001FB	AL060732 Drosophila

c 25	52.6	4.6	1086	9	CNS008YXK	AL096962 Drosophila
c 26	52.2	4.6	498	4	BJ396861	BJ396861 Drosophila
c 27	52.2	4.6	1101	9	CNS008EQ	AL069526 Drosophila
c 28	52	4.6	733	9	BX244451	BX244451 Danio rer
c 29	52	4.6	894	2	BF262513	BF262513 HV CEa000
c 30	51.4	4.5	575	8	BZ780846	BZ780846 ii20h11.9
c 31	51.4	4.5	1085	9	CNS0124K	AL101102 Drosophila
c 32	51.2	4.5	1101	9	CNS008YX	AL071206 Drosophila
c 33	50.6	4.5	400	9	CG807323	CG807323 1118080G0
c 34	50.6	4.5	430	9	CG803196	CG803196 1118039H0
c 35	50.6	4.5	528	4	BJ414423	BJ414423-BJ414423
c 36	50.6	4.5	1043	9	CNS0145P	AL103735 Drosophila
c 37	50.6	4.5	1193	9	CG745316	CG745316 P038-1-G0
c 38	50.4	4.4	677	9	DR61N5S	AL987387 Danio rer
c 39	50.4	4.4	1974	3	CR734974	CR734974 Tetraodon
c 40	50.2	4.4	479	4	BJ415760	BJ415760 BJ415760
c 41	50.2	4.4	855	9	CNS04P5D	AL300874 Tetraodon
c 42	50.2	4.4	928	9	CNS000DKY	AL071865 Drosophila
c 43	50.2	4.4	945	9	CNS04D0K	AL285149 Tetraodon
c 44	50.2	4.4	1225	9	CNS0161D	AL106171 Drosophila
c 45	50	4.4	616	4	BJ417643	BJ417643 BJ417643

## ALIGNMENTS

RESULT 1	CNS0039G	1101 bp	DNA	linear	GSS 03-JUN-1999
LOCUS	CNS0039G	Drosophila melanogaster genome survey sequence	TET3 end of BAC #		
DEFINITION	AL063921	BAC08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.			
ACCESSION	AL063921	GSS.			
VERSION	AL063921.1	GI:4941778			
KEYWORDS		Drosophila melanogaster (fruit fly)			
SOURCE		Drosophila melanogaster			
ORGANISM		Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
REFERENCE		1 (bases 1 to 1101)			
AUTHORS		Genoscope.			
TITLE		Direct Submission			
JOURNAL		Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)			
COMMENT		- Web : www.genoscope.cns.fr			
		Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoss in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila_bac.htm">http://bacpac.med.buffalo.edu/drosophila_bac.htm</a> .			
FEATURES		Location/Qualifiers			
source		1..1101			
		/organism="Drosophila melanogaster"			
		/mol_type="genomic DNA"			
		/db_xref="taxon:7227"			
		/clone_lib="BAC08K10"			
		/note="end : TET3"			
ORIGIN					
Query Match		5.8%	Score 66;	DB 9;	Length 1101;
Best Local Similarity		19.9%	Pred. NO. 0.00038;		

	Matches	138;	Conservative	271;	Mismatches	277;	Indels	:8;	Gaps	2;
Qy	391	ACAGAAATCCTTTGATT	TTTTAAAAACAATAATTAAGAAGTCTA	TATTATTCGTGAACAAGCT	450					
Db	398	ATAWAWWWMTTTTTT	TAWAAAABAAAATAAT	TWAAWAWAAAAAAT	TWAAAABAWAAAW	457				
Qy	451	TTTGACGCCAGTTT	TACAGAAAGTCGTCCTATTTT	AACAGAGTTCATAT	TGTAATTAAGT	510				
Db	458	AWTAWMTTWTATW	AAAAAATAAATTTT	TTTWTATATWMTT	TTTTWWWTTTAAAW	517				
Qy	511	AATCATAAAGATTTT	TAAGCAGTAGCGCACTC	CTCATCTGATGAGCCAA--	--CGTTTA	567				
Db	518	AAAAAAAAAAAAA	AAWAATAATTWTTT	TTTWTAAWATAAACMAWYHYHTT	577					
Qy	568	ATCACTTTGGCAAT	CATTGACGAGATTTGATGGTAGT	CTTCTCAAGTAATCTT	GAGA	627				
Db	578	YTTHYYUVTW	TMTHYHMYTHAWAHT	TTWHYTYAMHHMHTWMTWAAWMHTTYTA	637					
Qy	628	GAATTTTTCAGCAGT	ATTACAGATGATATTGAGACCGT	TGAGGTGATTTTCTCACCAAGC	687					
Db	638	AYYYYYCMYYYH	MMHNHAANAANWT	TTWTHAYHWATYHYYYMCMHMCHT	697					
Qy	688	CAAATCTTGTTCAGA	AGTGAACAATTTCTTTTATACACGCCT	CTCTTAGAAGGAAATVAT	747					
Db	698	CHHCYYUHTYAT	JNHTTHHWAHYUMWYUMWYA	UYMWYUCTACTYVYHH----	HNYHWA	752				
Qy	748	CCCATACAGACCGT	TTATTAAATCACAGAAATTT	GAGACGGGTGTTTCAATACC	807					
Db	753	YHTTWYAWAHAM	MMNNHANVAALAAWALT	HTTHYNHTTHVMHNTYMHYNYTCCYMCST	812					
Qy	808	TCCCTTCGCCACG	CTATGGAACGTGCTCTTGTGATTT	CTAATGCTACTCAAAATGGTACT	867					
Db	813	YHCWYUHTYAT	CTWTHHNMWTTWTHWYHNT	TMHNTTTHWAWHHTTWCWWWHAATTW	872					
Qy	868	GTTAAGCTTCGAGAT	TACTCAAAATCATATTT	CAGCTCATGTTAACCTCACCTGAGT	TGGT	927				
Db	873	ATHCWA	CMTHMNHMMHNNHNAACHANNHT	MMCCHHNMECTCHHNTYHNTCHMMWH	932					
Qy	928	AAGGTAAACGAGG	ATTTAGATATTTGTTAGTCAGT	CTGTGATTTAACTATCACGCTTC	987					
Db	933	MHWHHHMMWMA	TWMTTMTMMCMCHHNC	MMHMMYMCCHYUCTCHTHATTHYH	992					
Qy	988	AATCCAACTTACCT	TATTTAGTCTTTAAAGCTAT	TAAAGTGAACAGGTAA	AAATTCAT	1047				
Db	993	MCTCYHCT	NHTTYTAT	TWAAWTAAWTTATWMMHMMHWA	TTWMMHMMHWA	1052				
Qy	1048	TTCTTATCACCA	GTTTCAGCAATTCACCTCAAC	1081						
Db	1053	HNHTWYHHTCT	YUHNHTYHMMWAAWMMWNNHY	1086						

RESULT 2	
BZ296086/c	
LOCUS	603 bp DNA linear GSS J1-OCT-2002
DEFINITION	CGL1849.f11 Candida glabrata Random Genomic Library Candida glabrata genomic clone CGL1849, genomic survey sequence.
ACCESSION	BZ296086
VERSION	BZ296086
KEYWORDS	GSS.
SOURCE	BZ296086.1 GI:24437749
ORGANISM	Candida glabrata Candida glabrata Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida. 1 (bases 1 to 603)
REFERENCE	Wong,S., Farees,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H. Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast Candida glabrata
JOURNAL	Genome Biol. 4 (2), R10 (2003)
MEDLINE	22508158
PUBMED	12620120
COMMENT	Contact: Wong S Department of Genetics, Smurfit Institute Trinity College Dublin

```

FEATURES
. source
    Dublin 2, Ireland
    Tel: 353 1 6082319
    Fax: 353 1 6798558
    Email: swong@tcd.ie
    Class: plasmid ends.
        Location/Qualifiers
            1. .603
                /organism="Candida glabrata"
                /mol_type="genomic DNA"
                /strain="CBS 138"
                /db_xref="taxon:5478"
                /clone="CG1849"
                /clone_lib="Candida glabrata Random Genomic Library"

```

## ORIGIN

Query Match	5.6%	Score 63.6	DB 8	Length 603
Best Local Similarity	49.0%	Pred. No. 0.0012		
Matches 199	Conservative	0	Mismatches 204	Indels 3
			Gaps	1
QY	12	TTCAANTTAATCGCACATTATTTATTCATGCTTTAAATACAACTAAACGTCGTATTAGCAC	71	
Db	405	TTATATTTCTGATTCATGATATAATTTATAAAAAATTATAAAAAATTATAATATTATATAAA	346	
QY	72	TAAAAATGCCATTCTCTATCTTTTCATCAATATAAAAAATTGAAAGTCACCTCTCAGCAGGTAAC	131	
Db	345	TATAAAAATAATAATAATAATATTAGAAAAATAATTTATTAATAATTAATAAAATTATAAC	286	
QY	132	TTTAAACAGGGCTAAACGGTCAAATATCAATTGAAAAACACTATTCTCTGTAGTAATGAAAA	191	
Db	285	TAAATCATGATTAATTTGGTTATATAGAAACTAAAAAGTTCAATTTAAATTAATTTAAAAAAGA	226	
QY	192	TGCTGGTTTGCTAATTTACCTCTCCAGGAGCTATTTTATTAGAAGCTAGTTTTTTTTTATTAA	251	
Db	225	TATAAAAAATTTCATTTTAAATTAATAATAATAAAAAATTAGATTATTAATTTGTAATATA	166	
QY	252	TATTTATTCAGTTTCGCCAGATATTAGTATAAATGTTTAAAGAAATTGAAACAACCAAGT	311	
Db	165	TAGTATTTAAATTTATTTAAATTTAAATTTAATAATAATAATAATAATAATAATAATAATA	106	
QY	312	TGTTTTTAAACAGTGGTAAATCAGAGATTACCTTTAAAGGAAAGAGTCTGCACAGTATCC	371	
Db	105	TGAATAATAAGTATTGAAT---ATAATTTAATTTATTAAAAATTTTAATAATCATTAATC	49	
QY	372	TGCTCTCAAGAAGTATCAACAGAAATCCTTTGATTTTTAAAAACA	417	
Db	48	AAAGATTATTAGGTATAAAAAGTTATCGAATATCTAAATTTGAAAAAGA	3	

RESIST 3

BZ295031	614 bp	DNA	linear	GSS 31-OCT-2002
CG1245.f1	Candida glabrata	Random Genomic	Library	Candida glabrata
genomic clone CG1245,	genomic survey sequence.			
BZ295031				
BZ295031.1	GI:24436038			
GSS.				
Candida glabrata				
Candida glabrata				
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida. 1 (bases 1 to 614)				
Wong, S., Fares, M.A., Zimmermann, W., Butler, G. and Wolfe, K.H.				
Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast Candida glabrata				
Genome Biol. 4 (2), R10 (2003)				
22508158				
12620120				
Contact: Wong S				
Department of Genetics, Smurfit Institute				
Trinity College Dublin				
Dublin 2, Ireland				
Tel: 353 1 6082319				
Fax: 353 1 6798558				

Email: swong@tcd.ie

Class: plasmid ends.

Location/Qualifiers

## FEATURES

source

1. .614

/organism="Candida glabrata"

/mol\_type="genomic DNA"

/strain="CBS 138"

/db\_xref="taxon:5478"

/clone="CG1245"

/clone\_lib="Candida glabrata Random Genomic Library"

## ORIGIN

Query Match 5.6%; Score 63.6; DB 8; Length 614;

Best Local Similarity 49.0%; Pred. No. 0.0012;

Matches 199; Conservative 0; Mismatches 204; Indels 3; Gaps 1;

QY 12 TTCAATTAATCGCACATTAATTTATTCATGCTTTAAATCAAACTAAACGTCGCTATTAGCAC 71  
 Db |||||  
 QY 405 TTATATTTCTGATTCATGATATAATATTATTAATAAATTATAAATTAATTAATTAATAA 346  
 Db |||||  
 QY 72 TAAATAGCCATCTCTATTTCTTTCATCAATATAAATTTGAAGTCATCTTACAGGAGTAAC 131  
 Db |||||  
 QY 345 TATAAATAATAAATAATATATATAGAAAATTAATTTAATAATAATAAATAATAAATATAAC 286  
 Db |||||  
 QY 132 TTTACAGGCTCAACGGTCAATATCAATTCGAACACATATTCCTGTAAGTAATGAAGA 191  
 Db |||||  
 QY 285 TAAATCATGATTAATTTGGTTATATAGAACTAAAGTTCATTTAAATTAATTAATAAAGA 226  
 Db |||||  
 QY 192 TCTCGTTTGTAAATACCTCTCCAGGAGCTATTTTATAGAGCTAGTTTTTTTATTA 251  
 Db |||||  
 QY 225 TAATAAATTTATTCATTTTAAATTAATTAATTAATAAATTTAGATTATTTGTAATATA 166  
 Db |||||  
 QY 252 TATTATTTCAAGTTTCCGAGATATTAGTATAAATGTTAAAGAAATGGAACACACCAAGT 311  
 Db |||||  
 QY 165 TAGTATTAATTAATTTAATAATTAATTAATAATAATAATTAATAATTAATAATACTATT 106  
 Db |||||  
 QY 312 TCTTTTAAACCGTGTAAATCAGATTTACCTTTAAAGGAAAGATGTTGACAGATATCC 371  
 Db |||||  
 QY 105 TGATAATAAAGTATTGGAAT---ATATTATTAATTTTAAATTTAATAATCATTAATC 49  
 Db |||||  
 QY 372 TCGTCTCAAGAAGTATCAACAGAAATCCTTTGATTTTAAAGAA 417  
 Db |||||  
 QY 48 AAGATTATTAGGTATAAAGAGTTATGATATCTAATTTGAAAGAA 3

## RESULT 4

CNS00EVL

LOCUS

DEFINITION

Drosophila melanogaster 1101 bp DNA linear GSS 04-JUN-1999  
 Drosophila melanogaster genome survey sequence T7 end of BAC:  
 BAC29B23 of RPCI-98 library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the Berkeley Drosophila Genome Project (BDGP).  
 The BDGP is constructing a physical map of the Drosophila  
 melanogaster genome using these BACs. For further information  
 please see <http://www.fruitfly.org> The BDGP Drosophila  
 melanogaster BAC library was prepared by Kazuo Osoegawa and  
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of  
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo.

NY. The library is named RPCI-98 and was constructed by partial  
 EcoRI digestion of Drosophila DNA provided by the BDGP from the  
 isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
 p1 and EST libraries. A more detailed description of the library  
 and how to order individual BAC clones, the entire library, or  
 filters for hybridization from the BACPAC Resource Center can be  
 found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

source

Location/Qualifiers

1. .1101

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/db\_xref="taxon:7227"

/clone="BACR29B23"

/clone\_lib="RPCI-98"

/note="end : T7"

## ORIGIN

Query Match 5.1%; Score 57.8; DB 9; Length 1101;

Best Local Similarity 31.2%; Pred. No. 0.026;

Matches 164; Conservative 92; Mismatches 269; Indels 0; Gaps 0;

QY 4 ATTCAAATTTCAATTAATCGCACATTAATTTATTCATGCTTTAAATCAAACTAAACGTCGT 63  
 Db |||||  
 QY 564 ATTAATAATTTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 623  
 Db |||||  
 QY 64 ATTAGCAGTAAATGCGCATTCCTATTCCTTCATCAATAAATAATTTGAAGTCATCTCTACA 123  
 Db |||||  
 QY 624 WATAAATTTTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 683  
 Db |||||  
 QY 124 GGAGTAACTTTAAACAGGGTCTAAACGGTCAATATCAATTTGAAACACATATTCTCTGAAGT 183  
 Db |||||  
 QY 684 WAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 743  
 Db |||||  
 QY 184 AATGAAATGCTGGTTTGTCTAATTAACCTCTCCAGGAGCTATTTTATTAAGAAGCTAGTTT 243  
 Db |||||  
 QY 744 AATWAWATAATATWATAATATATWTTTAAWATWAAWATWAAWATWAAWATWAAWATAWA 803  
 Db |||||  
 QY 244 TTTATTAATTAATTTTCAAGTTTCCAGATATTAGTATAAATGTTTAAAGAAATTTGAACAA 303  
 Db |||||  
 QY 804 WATAAATATAWATAWATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 863  
 Db |||||  
 QY 304 CACCAAGTTGTTTAAACAGGTGTAATCAAGATTTACCTTAAAGGAAAGATGTTGAC 363  
 Db |||||  
 QY 864 WAAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 923  
 Db |||||  
 QY 364 CAGTATCTCTCTCAAGAAAGTATCAACAGAAATTCCTTTGATTTTAAATAAATAAATA 423  
 Db |||||  
 QY 924 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 983  
 Db |||||  
 QY 424 TTGAAGTCTATATTGCTGAAACAGCTTTTGCAGCCAGTTTACAGAAAGTCTCTATT 483  
 Db |||||  
 QY 984 TATWTATATATWTTAATWATAATATTTTAAWWTATATTTTAAWAAATAATATATATA 1043  
 Db |||||  
 QY 484 TTAACAGGATTCATATTGTTATTAAGTAAATCATTAAGATTTTAA 528  
 Db |||||  
 QY 1044 WWTATATATAWAAWAAATTAATTTATATATATAWATAWATAWATAWATAWATAWATA 1088  
 Db |||||

## RESULT 5

CNS0145P

LOCUS

DEFINITION

Drosophila melanogaster 1043 bp DNA linear GSS 26-JUL-1999  
 BACN1G11 of DrosBAC library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 1043)

ACCESSION	CG8031197
VERSION	CG8031197.1
KEYWORDS	GI:38238983
SOURCE	GSS
ORGANISM	Zea mays
	Zea mays
	Zea mays
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE	1 (bases 1 to 429)
AUTHORS	Walbot,V.
TITLE	Maize genomic sequences found using engineered RescueMu transposon

REFERENCE 1 (bases 1 to 1101)  
 LOCUS Genoscope.  
 DEFINITION Direct Submission  
 TITLE Submitted (23-JUN-1999) Genoscope - Centre National de Sequencage :  
 JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 COMMENT Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the Berkeley Drosophila Genome Project (BDGP).  
 The BDGP is constructing a physical map of the Drosophila  
 melanogaster genome using these BACs. For further information  
 please see <http://www.fruitfly.org> The BDGP Drosophila  
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and  
 Aaron Mamoser in Pieter de Jong's laboratory in the Department of  
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
 NY. The library is named RPCI-98 and was constructed by partial  
 EcoRI digestion of Drosophila DNA provided by the BDGP from the  
 isogenic strain Y2; cn bw sp, the same strain used for the BDGP's  
 p1 and EST libraries. A more detailed description of the library  
 and how to order individual BAC clones, the entire library, or  
 filters for hybridization from the BACPAC Resource Center can be  
 found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
 Location/Qualifiers

FEATURES  
 source  
 1..1101  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"  
 /clone="BACR29823"  
 /clone\_lib="RPCI-98"  
 /note="end : T7"

## ORIGIN

Query Match 5.0%; Score 56.2; DB 9; Length 1101;  
 Best Local Similarity 33.9%; Pred. No. 0.059;  
 Matches 167; Conservative 82; Mismatches 237; Indels 7; Gaps 1;  
 QY 10 TTTTCAATTAATCGCACATTATTTATTCATGCTTTAAATACAACTAAACGCTATTAGC 69  
 DB 927 TTTTCTTTTWTWTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 868  
 QY 70 ACTAAAAATGCCATCTTCTTCTCATCAATAAAATGAAAGTCACCTTACAGGAGTA 129  
 DB 867 ATTATATATTAATTTTWTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 808  
 QY 130 ACTTTAAGCGGCTTAACGGTCAATATCAATTTGAAACACATTTCTCGTAAAGTAAGAA 189  
 DB 807 TATTTTWTWTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 748  
 QY 190 AATGCTGGTCTCAATTAACCTCTCCAGGAGCTATTTTATTAAGAAGCTAGTTTATT 249  
 DB 747 WAATTTTWTWTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 695  
 QY 250 AATATTTATTTCAAGTTTGCAGATATTTAGTATAAATGTTTAAAGAAATGAAACACCAA 309  
 DB 694 TTTTWTATTTTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 635  
 QY 310 GTTGTTTTAAACGCTGTAATACAGATTAACCTTAAAGGAAAGATGTTGACCAAGTAT 369  
 DB 634 AAAATTTTATWAAAAAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 575  
 QY 370 CCTCGCTACAGAGAGTATCAACAGAAATCCCTTTGATTTTAAACAAAATTTATTTGAAG 429  
 DB 574 AAATTTTWTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 515  
 QY 430 TCTATTTATTTGCTGAAACAGCTTTTGCAGCCAGTTTACAAAGAAAGTCGCTATTATTTAACA 489  
 DB 514 WAAAAATTTTWTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 455  
 QY 490 GGAGTTCAATTTG 502  
 DB 454 RATKDKKKKKKG 442

## RESULT 8

CNS0161D/c  
 LOCUS

DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC  
 BACN15C18 of DrosBAC library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.

ACCESSION AL106171 GI:5620504  
 VERSION  
 KEYWORDS  
 SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE 1 (bases 1 to 1225)  
 AUTHORS Genoscope.  
 TITLE Direct Submission

JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr  
 - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the European Drosophila Genome Project (EDGP) -  
<http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC  
 library (Dros BAC) was made by Alain Billaud at CEPH (Centre  
 d'Etude du Polymorphisme Humain) with funding provided by a MRC  
 project grant. The DNA was prepared from embryos by Alain Bucheton  
 and Genevieve Payan. It has been constructed in the vector  
 pBelobAC11.

FEATURES  
 source  
 1..1225  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"  
 /clone="BACN15C18"  
 /clone\_lib="DrosBAC"  
 /plasmid="pBelobAC11"  
 /note="end : SP6"

## ORIGIN

Query Match 5.0%; Score 56.2; DB 9; Length 1225;  
 Best Local Similarity 36.7%; Pred. No. 0.06;  
 Matches 106; Conservative 50; Mismatches 133; Indels 0; Gaps 0;  
 QY 237 TAGTTTTTTTATTAATTAATTTTCAAGTTTGCAGATTAATAGTATAAATGTTTAAAGAAAT 296  
 DB 1159 WATTTATTTATTTAAAWAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1100  
 QY 297 TGAACCAACCAAGTTGTTTAAACCAAGTGGTAAATCAGAGATTACCTTAAAAAGGAAAGA 356  
 DB 1099 AAWAAWAAWAAWATTTWTWKAATATAAARTAWATATATAAATWAAWAAWTTWAA 1040  
 QY 357 TGTTGACCATGATCCTCGCTACAGAAAGTATCAACAGAAATCCTTTGATTTTAAAAAC 416  
 DB 1039 AAATATATTTTTTTTTTTTAAWAAAAAATAAATAATTTATTTATTTATTTATTTATTT 980  
 QY 417 AAATATTATGAAGTCTATTTGCGAAACAGCTTTTGCAGCCAGTTTACAGAAAGTCG 476  
 DB 979 AAAAATAAATTTATTTWAAWATTTTTRAAAWTTTTTTTTTTAAWAAWTTTATTTT 920  
 QY 477 TCCTATTTTAAACAGAGTTTCATTTGTTATTAAGTAATCATAAAGATTTT 525  
 DB 919 TWTRTTWTTRRRARAATTTTATTTTATTTTATTTTATTTTATTTTATTTT 871

## RESULT 9

CNS00FYG/c

LOCUS CNS00FYG/c 1101 bp DNA linear GSS 03-JUN-1999  
 DEFINITION Drosophila melanogaster genome survey sequence TE13 end of BAC:  
 BACR32N04 of RPCI-98 library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.

ACCESSION AL071206  
 VERSION AL071206.1 GI:4951245  
 KEYWORDS  
 SOURCE Drosophila melanogaster (fruit fly)



ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osogawa and  
Aaron Mammosier in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
p1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
Web at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).  
Location/Qualifiers  
1..1101  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR32N04"  
/clone\_lib="RPCI-98"  
/note="end : TET3"

FEATURES  
source

ORIGIN  
Query Match 4.9%; Score 55.6; DB 9; Length 1101;  
Best Local Similarity 38.9%; Pred. No. 0.081;  
Matches 199; Conservative 49; Mismatches 258; Indels 5; Gaps 2;  
QY 225 TTTTATAGAGCTAGCTTTTATTATATATATTTCAAGTTTGCAGATATAGTATATAA 284  
DB 1040 TATATATATTTTA 981  
QY 285 TCGTTAAGAAATGGAACACCAAGTTGTTTTAACCAAGTGGTAAATCAGAGATTACCTT 344  
DB 980 WATATATATTTTAAAGAAATTTAAATTTTATATATATATATATATATATATAT 921  
QY 345 AAAAGGAAAGATGTTGACCGATCTCTGCTACAGAAAGTATCAACAGAAATCCCTT 404  
DB 920 WTAT 861  
QY 405 GATTTTAAAAACAAATTTATTTGAAGTCTATATTTCTGAAACAGCTTTTGCAGCC--AG 461  
DB 860 WATTWTATATATTTAAATATATATATATATATATATATATATATATATATATAT 801  
QY 462 TTTACAGAAAGTCCTCTATTTTAAAGGAGTTCATATGTTATTAAGTAAATCAATAAGA 521  
DB 800 TTWTATATTTAT 741  
QY 522 TTTTAAAGCAGTAGGAGCTGCTCTCATCGTATGAGCCACGTTTAAATCACTTTGGCAA 581  
DB 740 WATWTATTTTAT 681  
QY 582 TACTTTCAGCAGATTGATGGTAGTTCTTCCAGTAAATCTTTTGAGAGAAATTTTCAGCAGT 641  
DB 680 TATATATAATTTNNKGGGKTKBKBBSBNNNNNTNTNNNTATATATATATATATAT 621  
QY 642 ATTTACAGATGATTTGACCGGTGAGGATTTTCTCACCAGCCAAATCTTGTTCAG 701  
DB 620 ATATATTTAT--TACTTAGAAAAATGCATCTAAATTTTATAGCCTCTAAAAATTCCTCAA 563  
QY 702 AAGTGAACACATTTCTTTTATACAGGCTC 732

Db 562 AAATTAATATTTGTTAATTTGCTGTACAGCTC 532  
BX159596 738 bp DNA linear GSS 13-MAR-2003  
Danio rerio genomic clone DKEY-117F6, genomic survey sequence.  
RESULT 10  
BX159596  
LOCUS  
DEFINITION  
ACCESSION BX159596  
VERSION BX159596.1 GI:27991215  
KEYWORDS GSS.  
SOURCE  
ORGANISM Danio rerio (zebrafish)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
Cypriniformes; Cyprinidae; Danio.  
REFERENCE  
1 (bases 1 to 738)  
Humphray, S.J., Huckle, E. and Durham, J.L.  
Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome  
Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humqu@sanger.ac.uk unpublished  
This sequence was generated from the SP6 end of BAC 117F6. 117F6 is  
part of the Daniokey BAC Library created by R. Plasterk and N.V.  
Keygene. Further details:  
[http://www.sanger.ac.uk/Projects/D\\_rerio/](http://www.sanger.ac.uk/Projects/D_rerio/).  
Location/Qualifiers  
1..738  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="DKEY-117F6"  
/tissue type="Testis"  
/note="Vector pIndigoBAC-536"

ORIGIN  
Query Match 4.9%; Score 55.4; DB 9; Length 738;  
Best Local Similarity 48.9%; Pred. No. 0.084;  
Matches 149; Conservative 0; Mismatches 156; Indels 0; Gaps 0;  
QY 49 ACACTAACGTCGTATTAGCACTAAATGCGCATTCCTATTCTTTCATCATATAAAATT 108  
DB 215 ACATGCAATATTCACACAGAAATCAAAAGTGTATAACATTTACTTTTGGAAATCCCTGTT 274  
QY 109 GAAGTCACCTCTACAGAGTAACCTTTAACGGGTCTAACGGTCAATATCAATATGAAAAAC 168  
DB 275 TTTAATATTATTAGACAATCTTTTATCAGGAGAGACGTATAGAAATATTTTAAAT 334  
QY 169 ACTATTCTGTGAAGTAAGAAATGCTGGTTTGTAAATACCTCTCCAGGAGCTATTATTA 228  
DB 335 ACAAAATCTTTTAAATATAAAATTTTAAAGAGATTCAATCAAAAGTAAATATTA 394  
QY 229 TTAGAAGCTAGTTTTTTTATATATATTTTCAAGTTTGCAGATATAGTATAAATGTT 288  
DB 395 TTAAGGTTTTAGATTGTTTAACTTATTGCTGTGTTAATGTCCTTTTATTTTATTTT 454  
QY 289 AAAGAAATGAACAACACCAAGTTGTTTAAACAGTCGTAAATCAGAGATTACCTTAAAA 348  
DB 455 TATTATTATTATAACAATTTTATTATATAAAATGTTGTAATTCGATTGTTTTCCTCA 514  
QY 349 GGAA 353  
DB 515 TGAAG 519  
RESULT 11  
CD387684  
LOCUS  
DEFINITION  
ACCESSION CD387684  
VERSION CD387684.1 GI:31224971  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 1036)  
NIH-MGC <http://mgc.mci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: Dr. Jamie Thompson, University of WI

cDNA Library Preparation: Gina Zastrow-Hayes

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: NDR#78 row: 9 column: 01

High quality sequence start: 6

High quality sequence stop: 439.

FEATURES  
source

1..1036  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/tissue\_type="embryonic trophoblasts, made from WA01 stem cells"  
/lab\_host="DH10B Tona"  
/clone\_lib="NIH MGC 173"  
/note="Vector: pDONR201; Site1: attP2; Site2: attP1;  
LIBR PRIMING - oligo dT; METHOD - full-length enriched;  
LIBR PROVIDER - Bradfield"

ORIGIN

Query Match 4.9%; Score 55.4; DB 6; Length 1036;  
Best Local Similarity 45.3%; Pred. No. 0.083;  
Matches 208; Conservative 0; Mismatches 244; Indels 7; Gaps 1;  
QY 49 ACAACTAAACGCTGCTATTAGCACTAAATAATGCTTCCTTCTTCATCAATAAAAAAT 108  
DB 354 ACACGTAAGTGAAGTCTCTTTAATAAAGGTTTGTTTCAAGTTAAAAAANAANA 413  
QY 109 GAAGTCACCTCTACAGGAGTAACTTTAAGAGGCTTAACAGGCTCAAAATATCAATGAAAC 168  
DB 414 AAAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA 473  
QY 169 ACTATTCCTGTAAGTAAGTAAGTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 228  
DB 474 TANAATAATGGAATTAATAAATAAAGCTTGAATTAATAAATTTTNTTNTNATATAATTT 533  
QY 229 TTAGAAGCTAGTTTTTTTATTATATATTTTCAAGTTTGCCAGATATTAGTATAAATGTT 288  
DB 534 NTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 593  
QY 289 AAGAAATTTGAACAACAACCAAGTTGTTTAAACAGTGGTAAATCAGAGATTACCTTAAAA 348  
DB 594 AAAAAAATAATATTTTATATATATATAAATAAATAAATAAATAAATAAATAAATAAATA 653  
QY 349 GGAAGAAGTGTGACAGGATCCTCGTCTACAGAGAGTATCAACAGAAATCCTTGATT 408  
DB 654 AATAAATAAATTTTATTA-----TATTAATAAATAAATAAATAAATAAATAAATAA 706  
QY 409 TTAATAAACAATAATTTGAAGTCTATTATTGCTGAAACAAGCTTTTGCAGCCAGTTTACAA 468  
DB 707 ATAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 766  
QY 469 GAAAGTCGCTTATTTTAAACAGAGGTCATATTGTATTA 507  
DB 767 AAAAAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 805

RESULT 12

CNS009G1/c  
LOCUS  
DEFINITION

Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR19J14 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.

ACCESSION

AL053529.1 GI:4935018

VERSION

GSS.

KEYWORDS

SOURCE

ORGANISM

Drosophila melanogaster (fruit fly)  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 876)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [seqrefgenoscope.cns.fr](mailto:seqrefgenoscope.cns.fr))  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Osogawa and  
Aaron Mammoss in Pieter de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp, the same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
Location/Qualifiers

1..876  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR19J14"  
/clone\_lib="RPCI-98"  
/note="end : TET3"

ORIGIN

Query Match 4.9%; Score 55.2; DB 9; Length 876;  
Best Local Similarity 37.4%; Pred. No. 0.096;  
Matches 185; Conservative 44; Mismatches 265; Indels 1; Gaps 1;  
QY 12 TTCAATTAATCGCACATTATTATTTCATGCTTTAATAACAACAACTAAGCTATTAGCAC 71  
DB 868 TANAWWATATNTAATAAAAAATATWATATATAATAATATATATATATATATANTA 809  
QY 72 TAAAAATGCCATTCCTATTCTTCATCAATAAAAAATGCAAGTCACCTCTACAGGAGTAAC 131  
DB 808 TATATATATTATTAATWATAAATAATANWAWATAAATTTWATAAATTTWATATATAWN 749  
QY 132 TTTAACAGGCTCAACGCTCAAAATCAATTTGAAAAACACTATTCCTCTGAAGTAATCAAAA 191  
DB 748 TATAATATATATWATNTATATATTTTNTATAAAAATATATATATATATATANAATATTT 689  
QY 192 TGCTGGTTGCTTAATACCTCTCCAGGAGCTATTTTATTAGAGCTAGTTTTTTTATTAA 251  
DB 688 WAAMWATATTTAAAAATATTTATNTAAWATATATATATATATATATATATATATTTAA 629  
QY 252 TATTATTTCAAGTTTGGCCAGATATTAGTATAAATGTTAAAGAAATTTGAACAACACCAAGT 311  
DB 628 TTTTATATTTTATATATAAATTTTATATAAATTTWATAAATTTWATTTWATATAATA 569  
QY 312 TGTTTTAAACCAAGTGGTAAATCAGAGATTACCTTAAAAAGGAAAGATGTTGACCAAGTATCC 371  
DB 568 TTATWTWTAATTTTTTTTWTATATATTTTTTTTWTATTTTTTTTATTTATTTTATAA 509  
QY 372 TCGTCTACAAGAGTATCAACAGAAAAATCCTTTGATTTTAAAAACAATAATTATTGAAGTC 431



```
ORIGIN
Query Match      4.8%; Score 54.8; DB 9; Length 1200;
Best Local Similarity 33.4%; Pred. No. 0.12;
Matches 115; Conservative 65; Mismatches 164; Indels 0; Gaps 0;
QY 129 AACTTTACAGGCTTACCGTCAAAATATCAATGAAACACTATCTCTGTAAGTAATGA 188
D 115 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 836 AAGTTTTTGGGTTTWWATAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAA 895
QY 189 AAATGCTGGTTGCTAAATCTCTCCAGGAGCTATTTTATAGAGCTAGTCTTTTAT 248
D 115 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 896 ADGAWTAATTTAACGAAAAATTTTAAWAAATTAATTTWAAWAAWAAWAAWAAWAAW 955
QY 249 TAAATATTATTTCAAGTTTCCAGATATTAGTATTAATGTTAAAGAAATTTGAACACCA 308
D 115 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 956 TTTAATTTTWTATTTTAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 1015
QY 309 AGTTGTTTACCAGGTGTAATCAGAGATACCTTTAAAGAAAGATGTTGACCACTA 368
D 115 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 1016 AAWTATTAAATATAAAATATAWTTTAAWAAWAAWTTTAAWAAWAAWAAWAAWAAW 1075
QY 369 TCCTCGTCTACAGAGATATCAACAGAAATCCTTTGATTTTAAACAAATATTGAA 428
D 115 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 1076 WAAAWWAAWAAWAAWAAWAAWTTTAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 1135
QY 429 GTCTATTATTGCTGAACAGCTTTTTCGACGCCAGTTTACAAGAAA 472
D 115 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 1136 ATTWTTTTWWAAWATTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1179

RESULT 15
BH941456
LOCUS BH941456
DEFINITION odg40C07.g1 B.oleracea002 Brassica oleracea genomic, genomic survey
sequence.
ACCESSION BH941456
VERSION BH941456.1 GI:23421516
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 705)
Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T.,
Nash,W., Rabinowicz,P.D. and Wilson,R.K.
Whole genome shotgun reads from Brassica oleracea
Unpublished (2002)
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Plate: odg40 row: c column: 07
Seq primer: -28RPpOT reverse
Class: shotgun
High quality sequence start: 16
High quality sequence stop: 551.
High quality sequence stop: 551.
Location/Qualifiers
1..705
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
(note="Vector: pOTW13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea T0100DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center.")

FEATURES
source
1..705
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/db_xref="taxon:3712"
/clone_lib="B.oleracea002"
(note="Vector: pOTW13; Whole genome shotgun library from
flowering buds. DNA was purified from a crude nuclear
prep using Brassica oleracea T0100DH3 buds provided by
Thomas Osborn at the University of Wisconsin. Genomic
DNA was provided by Pablo Rabinowicz (CSHL) and the
shotgun library prepared at Washington University Genome
Sequencing Center.")

ORIGIN
Query Match      4.8%; Score 54.6; DB 8; Length 705;
Best Local Similarity 48.5%; Pred. No. 0.13;
Matches 180; Conservative 0; Mismatches 189; Indels 2; Gaps 1;
QY 160 ATTGAAAACACTATTCTGTAAAGTAATGAAATGCTGGTTTGCTTAATTAACCTCTCCAGGA 219
D 115 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 36 ATAGCAATCAACCCACCCCTGAAGCGTCTCAAGTCTTAGTTTTTAAATTTTACTTTACAAT 95
QY 220 GCTATTTTATTAGAAGCT--AGTTTTTTTATTAATATTTTCAAGTTTGGCAGATATTA 277
D 115 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 96 TCCGCTTTTGTGAATTTGAAATGTTCTCAGTCTCTATATTTTAAATTAACGGATATTC 155
QY 278 GTATAAATGTTAAAGAAATGGAACAACACCAAGTTGTTTAAACCAAGTGGTAAATCAGAGA 337
D 115 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 156 ACTTAAATGTCAAAAAACCTTAGTCACTCGAACTTTGGTTATACCGTAGTTTCTTGCAAT 215
QY 338 TTACCTTTAAAGGAAAGATGTTGCACAGATATCTCGTCTACAAGAGTATCAACAGAAA 397
D 115 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 216 GCAAGTTTAAATCATAGAGTGAATACCGCAACATACGTCGGCACCGAACACATGAGCCC 275
QY 398 ATCCTTTGATTTTAAAAACAAAATTTATGAAAGTCTATTATTGCTGAAACAGCTTTTGAG 457
D 115 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 276 ATAATAATCATATATATATATATAATAAATTTTATTTTGGAAAATTTGGCTTTCAA 335
QY 458 CCAGTTTACAAGAAGTCTGCTCTATTTTAAACAGAGTTCATATTGTTAATAGTAAATCATA 517
D 115 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 336 TTAACCTGCAAGAAAAAATAAAACGTTTACAAGCCCGTTGAGCTTAAAGTTAAAAATACA 395
QY 518 AAGATTTTAAA 528
D 115 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 396 ACGTTTATAAA 406

RESULT 16
CNS0151M
LOCUS CNS0151M
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACN14L05 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL105604
VERSION AL105604.1 GI:5617618
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1061)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequenage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
Location/Qualifiers
1..1061
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone_lib="BACN14L05"
/clone_lib="DrosBAC"
/plasmid="pBelOBAC11"
/note="end : T7"

FEATURES
source
1..1061
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone_lib="BACN14L05"
/clone_lib="DrosBAC"
/plasmid="pBelOBAC11"
/note="end : T7"

ORIGIN
```

Query Match 4.8%; Score 54.4; DB 9; Length 1061;  
 Best Local Similarity 35.6%; Pred. No. 0.15;  
 Matches 143; Conservative 65; Mismatches 192; Indels 2; Gaps 1;

QY 243 TTTTATTAATATTTCAAGTTGCGAGATATTAGTATAAATGTTAAAGAAATTCAGCA 302  
 DB 654 TTTTATTAATATTTCAAGTTGCGAGATATTAGTATAAATGTTAAAGAAATTTMAWT 713

QY 303 ACACCAAGTTGTTTAAACAGTGGTAAATCAGAGATTACCTTAAAGAGAAAGATGTTGA 362  
 DB 714 AATTTTAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAA 773

QY 363 CAGATATCTCGTCTCAAGAGATATCAACAGAAATCTTTGATTTTAAAT - ACAAAA 420  
 DB 774 ATAATTTTAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAA 833

QY 421 TTTTGAAGTCTATTTCTGCAACAGCTTTTGCAGCAGTTTACAGAAAGTCTGCT 480  
 DB 834 AATTTTAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAA 893

QY 481 ATTTTAAACAGGATTCATATTTGATTAATCAATCAATCAATCAATCAATCAATCAATCA 540  
 DB 894 AATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAA 953

QY 541 GACTCTCATGATGAGCAAGTTTAAATCACTTTGAGCAATACATTCAGCAGATTTGATG 600  
 DB 954 TTTTATTAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAA 1013

QY 601 GTAGTTCTTCCAGTAATCTTTGAGAGAAATTTTCAAGCAGTA 642  
 DB 1014 AATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAA 1055

RESULT 17  
 CNS014PQ/c  
 LOCUS  
 DEFINITION  
 Drosophila melanogaster genome survey sequence Sp6 end of BAC  
 BAC1222 of Drosophila library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Drosophila melanogaster (fruit fly)  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
 1 (bases 1 to 987)  
 Genoscope.  
 Direct Submission  
 Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT  
 Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaut at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.

FEATURES  
 Location/Qualifiers  
 1..987  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"  
 /clone="BACN12P22"  
 /clone\_lib="DrosBAC"  
 /plasmid="pBelOBAC11"  
 /note="end : Sp6"

ORIGIN  
 Query Match 4.8%; Score 54; DB 9; Length 987;

Best Local Similarity 32.0%; Pred. No. 0.18;  
 Matches 108; Conservative 78; Mismatches 148; Indels 4; Gaps 1;

QY 152 AATATCAATTCGAAACACTATTCCTGTAAGTAAGTAAATGCTGGTTGCTTAATACCT 211  
 DB 985 AATATCAATTCGAAACACTATTCCTGTAAGTAAGTAAATGCTGGTTGCTTAATACCT 926

QY 212 CTCCAGGAGCTATTTTATTAGAGCTAGTTTATTAATATTAATATTAATATTAATATTA 271  
 DB 925 TTAATTTTATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAA 866

QY 272 ATATTAGTATTAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAA 327  
 DB 865 TTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAA 806

QY 328 AATCAGAGATTCCTTTAAAGAGGAAAGATGTTGACCAAGTATCTCTGCTCAAGAAAGTA 387  
 DB 805 WAATTAATTAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAA 746

QY 388 TCAACAGAGAAATCTTTGATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAA 447  
 DB 745 WTAATTTTATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAA 686

QY 448 GCTTTGCGAGCTTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAA 485  
 DB 685 TTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAAAGAAATTTTAA 648

RESULT 18  
 CNS020K7/c  
 LOCUS  
 DEFINITION  
 Tetraodon nigroviridis genome survey sequence T7 end of clone  
 222L11 of library G from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Tetraodon nigroviridis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE  
 1  
 Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizeaux,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.  
 Estimation of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence  
 Nat. Genet. 25 (2), 235-238 (2000)

JOURNAL  
 MEDLINE  
 PUBMED  
 20296633  
 10835645

REFERENCE  
 2  
 Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizeaux,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.  
 Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis  
 Genome Res. 10 (7), 939-949 (2000)

JOURNAL  
 MEDLINE  
 PUBMED  
 20359837  
 10899143

REFERENCE  
 3 (bases 1 to 1092)  
 Genoscope.  
 Direct Submission  
 Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT  
 This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.

FEATURES  
 Location/Qualifiers  
 1..1092





estimate of human gene number provided by sequencing Tetraodon nigroviridis DNA sequence data. Genet. 25 (2): 235-238 (2000).

JOURNAL  
TITLE

MEDLINE	20296633
PUBMED	10835645
REFERENCE	2
AUTHORS	Roezt Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Surin,W., Bernot,A. and Weissenbach,J.
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL	Genome Res. 10 (7), 939-949 (2000)
MEDLINE	20359837
PUBMED	10899143
REFERENCE	3 (bases 1 to 1092)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, Please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES	Location/Qualifiers 1..1092 /organism="Tetraodon nigroviridis" /mol_type="genomic DNA" /db_xref="taxon:99883" /clone_"222L11" /clone_lib="G" /note="Genoscope sequence ID : COAG222CF06LP1-end : T7"
ORIGIN	
Query Match	4.7%; Score 53.4; DB 9; Length 1092;
Best Local Similarity	38.0%; Pred. No. 0.25;
Matches	163; Conservative 52; Mismatches 205; Indels 9; Gaps 1;
Qy	4 ATTCAATTTTCAAATAACGCACATTATTTATTCATCGCTTTAAATACAACATAAAGTGCTGT 63
Db	614 AWTTTATTTTAAWWAAAAAAAATTTTTTTTTTTTAWAAAAAAWAAAAWWWAAAAA 673
Qy	64 ATTAGCACTAAAAATGCCATTCCCTATTCTTTTCATCAATAAAAAATGAAGTCACCTCTACA 123
Db	674 TTAAAAAAAATTTTTTTTTTTTTTTTTTTTTTTTAAAAAAAATTAATTAATTAATAA 733
Qy	124 GGAGTAACTTTAACAGGGTCTAACGGTCAAATATCAATGAAAAACACTATTCCTGTGAAGT 183
Db	734 WAAAAAAWAAAAWAAAAWTTTWTTTTTTTTWAAAAWAAAAWAAAAWTTTATTTTAAW 793
Qy	184 AATGAAAATGCTGGTTTGCTAATTTACCTCCTCCAGAGCTATTTTATTAGAGCTAGTTT 243
Db	794 WATWAWTTTWTATTTTAAWAAA-----ANAWTTTATTTATTAATAAAAAWAAAAATA 844
Qy	244 TTTATTAATATTTATTTCAAAGTTTGCCAGATATTAGTATAAATGTTTAAAGAAATTTGAACAA 303
Db	845 TTTTAAAAAAWTTTAAWAAATTTWAATTTAAAAAAAATTAATAAAWAAWTTTWTTTT 904
Qy	304 CACCAAGTTGTTTAAACAGGTGGTAAATCAGAGATTACCTTAAAGGAAAGAAGTTGTGAC 363
Db	905 TTTTWTWAWTTWTAATTTTTTTTTTTTTTTTTTTWTWTWTAAWAAAAAAAATATA 964
Qy	364 CAGTATCCTCCTCTACAAGAAGTATCAACAGAAAATCCTTTGATTTTAAAAACAAAAATTA 423
Db	965 AAWNAAAWTTTAAATTTATTTTAAAAAAAATTAATAAATAAATAAATAAATAAATA 1024
Qy	424 TTGAAGTCT 432
Db	1025 TTTATATAT 1033

Accession	CL469446	GI:45872351
Version	CL469446.1	
Keywords	GSS.	
Source	Arabidopsis thaliana (thale cress)	
Organism	Arabidopsis thaliana	
Reference	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.	
Authors	1 (bases 1 to 969)	
Journal	Sessions, A., Burke, E., Presting, G., Aux, G., McElver, J., Patton, D., Dietrich, B., Ho, P., Bacwaden, J., Ko, C., Clarke, J. D., Cotton, D., Bullis, D., Snell, J., Miguel, T., Hutchison, D., Kimmerly, B., Mitze, T., Katagiri, F., Glazebrook, J., Law, M. and Goff, S. A.	
Pubmed	A high-throughput Arabidopsis reverse genetics system	
Comment	Plant Cell 14 (12), 2985-2994 (2002)	
Features	Class: TDNA tagged.	
Source	Location/Qualifiers	
Origin	1..969	
Query Match	4.7%; Score 53; DB 9; Length 969;	
Best Local Similarity	46.1%; Pred. No. 0.3;	
Matches	243; Conservative 0; Mismatches 276; Indels 8; Gaps 2;	
QY	1 ATGATTCAAATTTTCAATTAATCGCACATATTATTATTCATGCTTTAAATCAACTAAACGT 60	
DB	382 ATATATATATTTATTTTATATAAATTTATTTTATTTTATTTTATTTTATTTTATTTTAAAT 441	
QY	61 GCTATTAGCACTAAAAATGCCATTCCTATTCCTTCATCAATAAAAAATGAAGTCATCT 120	
DB	442 ATTTAAATTTATATAAAATTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTAT 501	
QY	121 ACAGGAGTAACTTTAACAGGCTCAAGGTCAAATATCAATTCGAAAACACTATTCCTGTA 180	
DB	502 ATAATATTAATAATATAATATNT- - - - - TTTTATTTTATTTATTTATTTTATTTATAT 558	
QY	181 AGTAATGAAATGCTGGTTGGCTAAATTAACCTCCAGGAGCTATTTTATTTAGAAAGCTAGT 240	
DB	559 ATAAAAATAATTTATATTTATTTATATAATAATAATAATAATAATAATAATAATAATAATA 618	
QY	241 TTTTATTTATTTATTTATTTTCAAGTTTGCCAGATATTAGTATAAATGTTAAAGAAATTTGAA 300	
DB	619 ATTTATAATTTATTTATTTTAAATA - - - - - AAAATATTTTATTTATTTATTTATTTATTT 673	
QY	301 CAACACCAAGTTGTTTTAACCAAGTGGTAATCAGAGATTAACCTTTAAAGGAAAGATGTT 360	
DB	674 ATATTATAATAATTTTAAAAAATAATTTTATAAAATATAAATTTAAAAAATATATTTTATTTA 733	
QY	361 GACCAGTATCTCGTCTACAGAAGTATCAACAGAAAAATCCCTTTGATTTTAAAAAACAATA 420	
DB	734 ATTTATATTTTATTTATTTATTTTAAATATATAAATAATAATAATAATAATAATAATAATA 793	
QY	421 TTATTGAAGTCTATTTATTTGCTGAAACAGCTTTTTTGCAGCCAGTTTTCACAGAAAGCTCGTCT 480	





## ORIGIN





```

QY 192 TCGTGGTTGCTAAATACCTCTCCAGGAGCTATTATTATTAAGAGCTAGTTTATTATTA 251
Db 583 NAKNNNAANANNAANNNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA
QY 252 TATTATTTCAGAGTTGCGAGATATTAGTATAAATGTTTAAAGAAATGAACAACACCAAGT 311
Db 643 TWTAAKTATAANKTKATAAATKTAAKTAAKTAAKTAAKTAAKTAAKTAAKTAAKTAAKTAAKT
QY 312 TGTTTTAAACAGTGGTAAATCAGAGATTACCTTAAAGAGAAAGATGTTTACCAAGTATCC 371
Db 703 KDKTKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
QY 372 TCGTCTACAGAAGTATCAACAGAAAATCCTTTGATTTTAAACCAAAAATTTATTGAATC 431
Db 763 KAAKTAAKTAAKAAATKTAATAAADAATKTCTTAAKTAAKTAAKTAAKTAAKTAAKTAAKT
QY 432 TATTATTTCGCGAAGCTTTGCGAGCCTTTTACAGAAAGTCGCTTATTTTAAACAGG 491
Db 823 TTDKAAKKKKAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG
QY 492 AGTTCAATATGTTAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATC 551
Db 883 AAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
QY 552 TATGAGCAAGCTTTTAACTCTTGGACATATCTTCAGCAGATTTGATGGTATGTTCTCC 611
Db 943 AAAGTGAKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK
QY 612 AAGTAAATCTTTGAGAGAAATTTT 634
Db 1003 KAAWAKAAKTAAADADDK 1025

```

## RESULT 32

```

CNS00FYG
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TE13 end of BAC:
BAC32N04 of RPI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL071206
VERSION
AL071206.1 GI:4951245
KEYWORDS
Drosophila melanogaster (fruit fly)
SOURCE
Drosophila melanogaster
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
AUTHORS
Genoscope.
TITLE
Direct Submission
JOURNAL
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammos in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"

```

## FEATURES

```

source
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"

```

```

/db_xref="taxon:7227"
/clone="BAC32N04"
/clone_lib="RPI-98"
/note="end : TE13"

```

## ORIGIN

```

Query Match
Best Local Similarity 4.5%; Score 51.2; DB 9; Length 1101;
Matches 157; Conservative 47; Mismatches 235; Indels 0; Gaps 0;
QY 90 TCTTTTCATCAATATAAAATTTGAAGTCACTTCTACAGGAGTAACTTTAAACAGGCTTAACGG 149
Db 618 TATATTATATATATATATANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
QY 150 TCAAAATCAATTTGAAAAACACTATTTCCTGTAAAGTAATGAAATGCTGTTTTCCTAAATAC 209
Db 678 ATATTTTATATATATATATATATATATATATATATATATATATATATATATATATATAT
QY 210 CTCCTCCAGGAGCTATTTTATTAAGAGCTAGTTTATTAATTAATTAATTAATTAATTAAT 269
Db 738 ATWTATATATTTTATTAANNAATATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT
QY 270 AGATATTAGTATAAATGTTAAAGAAATTTGAACAACACCAAGTTGTTTAAACCAGTGGTAA 329
Db 798 WAATWAAATATATAWMAAAATTAATAAAATTAATAAAATTAATAAAATTAATAAAATTAATA
QY 330 ATCAGAGATTACCTTTAAAGAAAGATGTTGACCGATCTCCTCGTCTACAAGAAGTATC 389
Db 858 ATWAAATATATTTTAAAAAATTTATATATATATATATATATATATATATATATATATATA
QY 390 AACAGAAAATCCTTTGATTTTAAAAACAAATTTATTGAAGTCTTATTCTCTGAAACAGC 449
Db 918 TAWAAATATATATATATATATATATATATATATATATATATATATATATATATATATAT
QY 450 TTTTCAGCCAGTTTCAAGAAAGTCTCTCTATTTTAAACAGGAGTTTCATATTTGTTTAAAG 509
Db 978 ATTTTATATATATATATATATATATATATATATATATATATATATATATATATATATAT
QY 510 TAAATCAATAAGATTTTAAA 528
Db 1038 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT

```

## RESULT 33

```

CG807323/c
LOCUS
DEFINITION
CG807323 400 bp DNA linear GSS 10-NOV-2003
survey sequence.
ACCESSION
CG807323
VERSION
CG807323.1 GI:38244959
KEYWORDS
GSS.
SOURCE
Zea mays
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
1 (bases 1 to 400)
AUTHORS
Walbot,V.
TITLE
Maize genomic sequences found using engineered RescueMu transposon
COMMENT
Unpublished (2001)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 1118080 column:3
Class: transposon-tagged.
Location/Qualifiers
1..400
/organism="Zea mays"
/mol_type="genomic DNA"

```



## source

```
1..528
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddv19106"
/sex="mat A"
/dev_stage="Growth phase"
/clone_lib="Dictyostelium discoideum cDNA library, VF"
```

## ORIGIN

```
Query Match 4.5%; Score 50.6; DB 4; Length 528;
Best Local Similarity 46.6%; Pred. No. 0.95;
Matches 158; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

QY 23 GCACATATTATTATTCATGCTTTAAATCACTAAACGTCGTATATAGCACTAAATATGCCA 82
Db 2 GCATATATTNATTCATACATTTATNAAACAATGAAATTCCTCGTAATATCATTTTCAT 61

QY 83 TTCCTATCTTTCATCAATAAATAATGAAGTCACCTTCTACAGGAGTAACCTTTTAAACAGGCT 142
Db 62 TAGTTCGTCGTATTCACTTCAAGAGTGTAAATCGCTTCACATGCTTCCACATCAA 121

QY 143 CTAACGCTCAAAATCAATGAAACACATATTCCTGTAAGTAATGAATAATGCTGTTTCC 202
Db 122 CTGCTTTAAAGATTGAAATGAAATATACAGATAAAATTTCAATAAAGATGTTTAG 181

QY 203 TAATTACCTCTCAGGAGCTATTATTTATTAGAAGCTAGTTTATTATTAATATATTCAA 262
Db 182 TTGAATGTTTACATGCACCAATTAATCCAGCAGATTAATATTTTCAAGTACTATG 241

QY 263 GTTTGCCAGATATTAGTATTAATGTTAAAGAAATGAACAACCAAGTTGTTTAAACCA 322
Db 242 GTCAAAATGTTCAAGTTGAGGTGTAGCAGGTATGAAGGTGTGCTGTTGTTTAAAGG 301

QY 323 GTGGTAATCAGAGATTACTTTAAAGAGAAAGATGTTG 361
Db 302 TTGGTAGTGGTGTACAGATTAAAGAGAAATGTTAG 340
```

## RESULT 36

## CNS0145P/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## JOURNAL

## COMMENT

```
CNS0145P
Drosophila melanogaster genome survey sequence 17 end of BAC
BACN1G11 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL103735
GI:5615346
GSS.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1043)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
Location/Qualifiers
1..1043
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
```

## FEATURES

## source

```
1..1043
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
```

## ORIGIN

```
Query Match 4.5%; Score 50.6; DB 9; Length 1043;
Best Local Similarity 35.3%; Pred. No. 1;
Matches 85; Conservative 49; Mismatches 107; Indels 0; Gaps 0;

QY 198 TTTCGTAATACCTCTCCAGGAGCTATTTTATTAGAAGCTAGTTTATTATTAATATAT 257
Db 854 TTTTATATTTTATATATATATATATATATATATATATATATATATATATATATATAT 795

QY 258 TTCAAGTTTGCAGATATTAGTATATAATGTTTAAAGAAATGCAACAACCAAGTTGTTT 317
Db 794 WWWTTTWTWTATTTTAAAGAAATGTTTAAAGAAATGTTTAAAGAAATGTTTAAAGAA 735

QY 318 AACCAAGTGAATCAGAGATTACCTTAAAGAGAAAGATGTTTGAACCATCTCTCGTCT 377
Db 734 TATAAATTTAAAGAAATGTTTAAAGAAATGTTTAAAGAAATGTTTAAAGAAATGTTT 675

QY 378 ACAAGAGTATCAACAGAAATCCTTGAATTTTAAAGAAATGTTTGAACCATCTCTCGTCT 437
Db 674 WAWAARAAWAWAWAARAAATATATTTTAAAGAAATGTTTAAAGAAATGTTTAAAGAA 615

QY 438 T 438
Db 614 T 614
```

## RESULT 37

## CG745316/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## COMMENT

```
CG745316 1193 bp DNA linear GSS 24-OCT-2003
P038-1-G09-ya Ppa EcoRI BAC Library Pristionchus pacificus genomic,
genomic survey sequence.
CG745316
CG745316.1 GI:37966242
GSS.
Pristionchus pacificus
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
1 (bases 1 to 1193)
Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
Buntjer,J., van der Meulen,M. and Sommer,R.J.
An integrated physical and genetic map of the nematode Pristionchus
pacificus
Mol. Genet. Genomics 269 (5), 715-722 (2003)
22835951
12884007
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
```

## FEATURES

## source

```
1..1193
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
(note="The library was generated by a partial digest of
the genomic DNA with EcoRI and cloning into the BAC
vector.")
```

## ORIGIN

```
Query Match 4.5%; Score 50.6; DB 9; Length 1193;
Best Local Similarity 45.7%; Pred. No. 1.1;
```









```

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38117
; LENGTH: 1137
; TYPE: DNA
; ORGANISM: Streptococcus pyogenes
US-10-282-122A-38117

Query Match          99.3%; Score 1126; DB 16; Length 1137;
Best Local Similarity 99.6%; Pred. No. 8.8e-225;
Matches 1129; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATGATTCATATTTCAATTAATCGCACATATTTATTCATGCTTTAAATACAACTAAACGT 60
   |||||
Db 1 ATGATTCATATTTCAATTAATCGCACATATTTATTCATGCTTTAAATACAACTAAACGT 60
   |||||

QY 61 GCTATTAGCACTAAATGCAATTCCTATTTCTTCTCATTAATAAAATCGAAGTCATCTT 120
   |||||
Db 61 GCTATTAGCACTAAATGCAATTCCTATTTCTTCTCATTAATAAAATCGAAGTCATCTT 120
   |||||

QY 121 ACAGGAGTAACTTTAAGAGGCTCTAAGCGTCAAAATATCAATTCGAAACACTATTCCTGTA 180
   |||||
Db 121 ACAGGAGTAACTTTAAGAGGCTCTAAGCGTCAAAATATCAATTCGAAACACTATTCCTGTA 180
   |||||

QY 181 AGTAATGAAATGCTGGTTGCTAATTAATCTCCAGGAGCTATTTTATAGAAGCTAGT 240
   |||||
Db 181 AGTAATGAAATGCTGGTTGCTAATTAATCTCCAGGAGCTATTTTATAGAAGCTAGT 240
   |||||

QY 241 TTTTATTAATAATATTTCAAGTTTGCAGATATTTAGTAAATGTTAAAGAAATGGA 300
   |||||
Db 241 TTTTATTAATAATATTTCAAGTTTGCAGATATTTAGTAAATGTTAAAGAAATGGA 300
   |||||

QY 301 CAACCAAGTTGTTTAAACAGTGTAAATCAGAGATTTACCTTAAAGGAAAGATGTT 360
   |||||
Db 301 CAACCAAGTTGTTTAAACAGTGTAAATCAGAGATTTACCTTAAAGGAAAGATGTT 360
   |||||

QY 361 GACCAATATCTCTGCTCAAGAGATATCAACAGAAATCTTTGATTTTAAACAAAAA 420
   |||||
Db 361 GACCAATATCTCTGCTCAAGAGATATCAACAGAAATCTTTGATTTTAAACAAAAA 420
   |||||

QY 421 TTATTAAGTCTATTAATGCTGAAACAGCTTTTGCAGCCAGTTTACAAGAAAGTCGTCT 480
   |||||
Db 421 TTATTAAGTCTATTAATGCTGAAACAGCTTTTGCAGCCAGTTTACAAGAAAGTCGTCT 480
   |||||

QY 481 ATTTTAAACAGGAGTTTCAATATGTTATTAAGTAATCATAAAGATTTTAAAGCAGTAGCGACT 540
   |||||
Db 481 ATTTTAAACAGGAGTTTCAATATGTTATTAAGTAATCATAAAGATTTTAAAGCAGTAGCGACT 540
   |||||

QY 541 GACTCTCATCGTATGAGCAAGCTTTAATCACTTTGGACAATCTTCAGCAGATTTGATG 600
   |||||
Db 541 GACTCTCATCGTATGAGCAAGCTTTAATCACTTTGGACAATCTTCAGCAGATTTGATG 600
   |||||

QY 601 GTAGTCTTCCAGTAATCTTTGAGAAATTTTTCAGCAGATTTTACAGATGATTTGAG 660
   |||||
Db 601 GTAGTCTTCCAGTAATCTTTGAGAAATTTTTCAGCAGATTTTACAGATGATTTGAG 660
   |||||

QY 661 ACCGTTGAGGATATTTTCTCACCAGCAATCTTTGTCAGAGTGAACATTTCTTTT 720
   |||||
Db 661 ACCGTTGAGGATATTTTCTCACCAGCAATCTTTGTCAGAGTGAACATTTCTTTT 720
   |||||

QY 721 TATACACGCTCTTAGAAGGAAATTTATCCGATACAGACCGTTTATTAATGACAGATTT 780
   |||||
Db 721 TATACACGCTCTTAGAAGGAAATTTATCCGATACAGACCGTTTATTAATGACAGATTT 780
   |||||

QY 781 GAGCGGAGGTTGTTTCAATACCAATCTTCCTCGCCACCGTATGGAACGTGCTTCTTG 840
   |||||

```

RESULT 2

```

US-10-282-122A-35834
; Sequence 35834, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35834
; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Streptococcus mutans
US-10-282-122A-35834

```

```

Db 781 GAGACGAGGTTGTTTCAATACCAATCCCTTCGCCACGCTATGGAACGTCCTTCTTG 840
QY 841 ATTTCTTAATGCTACTCAAAATCGTACTGTTAAGCTTTGAGATTACTCAAAATCATATTTCA 900
   |||||
Db 841 ATTTCTTAATGCTACTCAAAATCGTACTGTTAAGCTTTGAGATTACTCAAAATCATATTTCA 900
   |||||

QY 901 GCTCATGTTAACTCACTGAGGTTGGTAAAGGTTAAACGAGGATTTAGATATTGTTAGTCAG 960
   |||||
Db 901 GCTCATGTTAACTCACTGAGGTTGGTAAAGGTTAAACGAGGATTTAGATATTGTTAGTCAG 960
   |||||

QY 961 TCTGTTAGTGAATTTAACTATCAGCTTCAATCCAACTTACCTTATTTAGATCTTTAAAGCT 1020
   |||||
Db 961 TCTGTTAGTGAATTTAACTATCAGCTTCAATCCAACTTACCTTATTTAGATCTTTAAAGCT 1020
   |||||

QY 1021 ATTTAAAGTGAAACAGTAGTAAAAATTCATTTCTTATATCAACAGTTCGACCAATTCACCTTAACA 1080
   |||||
Db 1021 ATTTAAAGTGAAACAGTAGTAAAAATTCATTTCTTATATCAACAGTTCGACCAATTCACCTTAACA 1080
   |||||

QY 1081 CCAGCGATGAGGAGGAAAGTTTATCCAAATTAATTTACCAAGTACGAAACAAAC 1134
   |||||
Db 1081 CCAGCGATGAGGAGGAAAGTTTATCCAAATTAATTTACCAAGTACGAAACAAAC 1134
   |||||

```

Query Match		60.9%; Score 690.8; DB 16; Length 1134;	
Best Local Similarity		75.6%; Pred. No. 5e-134;	
Matches 857; Conservative		0; Mismatches 277; Indels 0; Gaps 0;	
QY	1	ATGATTCAATTTTCAATTAATCGCACATTTATTTATTCATGCTTTTAAATACAACTAAACGT	60
DB	1	ATGATAAATTTTCAATTAATAAAGTTTTTCTTACAAGCCTTAAATGCTACCAAGCA	60
QY	61	GCTATTAGCACTAAAAATGCCATTCCTATTCTTTTCATCAATAAAAAATGAAGTCACCTCT	120
DB	61	GCTATTAGTTCTAAAAATGCTATTCTTCTTTCTAGTTTAAAAATGAAGTGAATCT	120
QY	121	ACAGGAGTAACCTTTAACAGGCTCTAACGGTCAATATCAATATGAATAACATATTCCTGTA	180
DB	121	CAGTCACTTACTTTAACAGGCTCTAATGGACAAATTTCTATTGAAAAATACGATTTACGT	180
QY	181	AGTAATGAAATGCTGGTTGCTTAATACCTCTCCAGGAGCTATTTTATTAGAGCTAGT	240
DB	181	GAAGAAGAAATGCTGGACTATTGGTTACTTCTCAGGAGCTATCTTACTTTGAAGCTAAT	240
QY	241	TTTTTTTATTAATATTATTTCAAGTTTGGCAGATATTAGTATATAAATGTTTAAAGAAATTGAA	300
DB	241	TTCTTTTATTAATATTGTTTCAAGCCTACTGATATCACTTTAGATTTTGAAGAAATTGAA	300
QY	301	CAACACAAAGTTGTTTAAACAGTGGTAAATCAGAGATTAACCTTAAAGGAAAAAGATGTT	360
DB	301	CAACATCAAGTTGTTTAAATAGTGGAAATCAGAAATTAACATTAAGGTAAGATGTG	360
QY	361	GACCAATATCCTCGTTGAGAAAGTATCAACAGAAATCTCTTGATTTTAAACAAAA	420
DB	361	GAACAAATATCCTCGTTGAGAAAGTGGAAACAAATATCTTTAAATTTTGAACAAAA	420
QY	421	TTATTGAAGTCTATTATTGCTGAAACAGCTTTTGAGCCAGTTTACAAAGAAAGTCGCT	480
DB	421	TTGTTAAAAACAATTTTCTGAAACAGCTTTTGCTAGTACCAGAAAGTCGACCA	480
QY	481	ATTTTAAACAGGAGTTTCAATTTGTTTAAATCAATAAAGATTTTAAAGCAGTAGCGACT	540
DB	481	ATTTTAAACAGGTTTCAATTTGTTTAAATCAATAAAGATTTTAAAGCAGTTGCTACA	540
QY	541	GACTCTCATGTTATGAGCCAAAGTTTATCACTTTGGACATACTTACGAGATTTTGATG	600
DB	541	GACTCTCATGCTAGTACGAAAGTTTAACTCTTTGACCAATCATCAGATGATTTTGAT	600
QY	601	GTAGTTCTTCCAAAGTAAATCTTTGAGAGAAATTTTTCAGCAGTATTTACAGATGATTTAG	660
DB	601	GTGCTCATTTCCAAAGCGGTTCTTTAGTGAATTTGAGCAGTATTTTACAGATGATTTGAA	660
QY	661	ACGGTTAGGATTTTCTCACCAGCCAAATCTTTGTTTCAAGAGTGAACATTTCTTTT	720
DB	661	TCGTGTTAGGATTTTCTCTTAAACAGTCAAAATCCTTTTGAAGTGAATATATTAGCTTC	720
QY	721	TATACAGGCTCTTAGAGGAAATTTATCCGNATACAGACCGTTTATTAATGACAGAAATTT	780
DB	721	TATACGCGCTCTTAGAGGAAATTTATCCTGATCTGATCTGCTTTGTTGGGCAATAACTTT	780
QY	781	GAGACGAGGTTGTTTCAATACCAATCCCTTTGCGCACGCTATGGAACGTCCTCTTG	840
DB	781	GAACAGAAAGTTGCTTCTCAATACTAATGCTCTTCTGTTATGCTATGAGAGCGTCAATTA	840
QY	841	ATTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTTGAGATTTACTCAAAATCATATTCA	900
DB	841	ATCTCAAAATGCTACTCAAAATGGTACTGTTTAAATTTGGAGATTTATCAACATCAAGTAA	900
QY	901	GCTCATGTTAACTCACTGAGGTTGGTAAGGTAACGAGGATTTTAGATATTGTTAGTCAG	960
DB	901	GCACATGTTAAATTCACCTGAGGTTGGAAAAAGTCAATGAAGAATCTTGATTTTGAAGCTTA	960
QY	961	TCGTGGTAGTTTAACTATCACTTCAATCCAGCTTACCTTATTTAGCTCTTTTAAAGCT	1020
DB	961	TCAGGAAATGATTTAAATCAATAGCTTTTAAACCACTTTACTTTGATCGAAGCTTTTAAAGCT	1020
QY	1021	ATTAAAGTGAAACAGTAAAAATTCATTTCTTATCACCAGTTTCGACCACTTCACCTTAACA	1080
DB	1021	CTTAAAGTGAGACTGTGCACAAATTCGTTTATTTTCCCAATTCGCCCTTTCACTTTTAA	1080
QY	1081	CCAGGCGATGAGGAAGATTTTATCCAAATTAATTAACACAGTACCAACAAAC	1134
DB	1081	CCGAGTGATACAGTGAAGAACTTCATTCATTTGATTACACTGTTTCGACAAAC	1134
RESULT 3			
US-08-961-527-20			
; Sequence 20, Application US/08961527			
; Publication No. US20020032323A1			
; GENERAL INFORMATION:			
; APPLICANT: Charles Kunsch			
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequence			
; NUMBER OF SEQUENCES: 391			
; CORRESPONDENCE ADDRESS:			
; ADDRESSES: Human Genome Sciences, Inc.			
; STREET: 9410 Key West Avenue			
; CITY: Rockville			
; STATE: Maryland			
; COUNTRY: USA			
; ZIP: 20850			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage			
; COMPUTER: HP Vectra 486/33			
; OPERATING SYSTEM: MSDOS version 6.2			
; SOFTWARE: ASCII Text			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/961.527			
; FILING DATE:			
; CLASSIFICATION: 424			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER:			
; FILING DATE:			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Brookes, A. Anders			
; REGISTRATION NUMBER: 36,373			
; REFERENCE/DOCKET NUMBER: PB340P1			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (301) 309-8504			
; TELEFAX: (301) 309-8512			
; INFORMATION FOR SEQ ID NO: 20:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 21338 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: double			
; TOPOLOGY: linear			
US-08-961-527-20			
Query Match			
Best Local Similarity			
Matches 805; Conservative			
Score 608.2; DB 8; Length 21338;			
Pred. No. 2.9e-116;			
Mismatches 328; Indels 0; Gaps 0;			
QY	1	ATGATTCAATTTTCAATTAATCGCACATTTATTTATTCATGCTTTTAAATACAACTAAACGT	60
DB	7076	ATGATTCAATTTTCAATTAATTAATAAATTTATTTCTACAAGCATTTAAATACTACTAAGAGA	7135
QY	61	GCTATTAGCACTAAAAATGCCATTCCTATTCTTTTCATCAATAAAAAATGAAGTCACCTCT	120
DB	7136	GCTATTAGTTCTTAAAAATGCCATTCCTATTCTTTATCAACAGTAAAAATTTGACGTGACCAAT	7195
QY	121	ACAGGAGTAACCTTTTAAACAGGCTTAAACGGTCAAAATATCAATTTGAAAAACACTATTTCCTGTA	180
DB	7196	GAAGTATTACTTTTAAATTTGTTTCAAAATGCTCAAAATTTCAATTTGAAATTTTATTTCTCAA	7255
QY	181	AGTAATGAAATGCTGGTTTCTGCTAAATTTACCTCTCCAGGAGCTATTTATTAGAACTAGT	240
DB	7256	AAAAATGAAGTCTGGTTGTTGTTAACTTTCTTTAGTTTCGATCTCTTCTTGAAGCTTCT	7315
QY	241	TTTTTTTATTAATATTATTTCAGTTTTCGAGATTTAGTATAAATGTTTAAAGAAATTTGAA	300
DB			

```
QY 301 CAACACCAAGTGTGTTTAAACAGTGGTAATCAGAGATTACCTTTAAAGGAAAGATGTT 360
Db |||
QY 7376 CAAAATCAAAATGTTTAAACAGTGGCAATCAGAAATACCTTAAAGGAAAGATAGC 7435
Db |||
QY 361 GACAGTATCTCGTCTACAAGAGTATCAACAGAAATCCTTTGATTTTAAACAAAAA 420
Db |||
QY 7436 GAACAAATATCCAGAAATCCAAGAAATTTACAGACGACTCTTTTAATACCTTGAACAAA 7495
Db |||
QY 421 TTATTGAAGTCTATTATGCTGAACAGCTTTTGCAGCCAGTTTACAAGAAAGTCTCT 480
Db |||
QY 7496 TTACTCAGAAATTAATGAACAGCTTTTGCAGTATACAAAGAGATCGTCCG 7555
Db |||
QY 481 ATTTTAAACAGGAGTTCATATTTGATTAAGTAATCATATAAGATTTTAAAGCAGTAGCGACT 540
Db |||
QY 7556 ATTTTAAACAGGAGTTCATATTTGATTAAGTAATCATATAAGATTTTAAAGCAGTAGCGACT 540
Db |||
QY 541 GACTCTCATCGTATGAGCAACGCTTTTAAACAGGAGTTCATATAAGATTTTAAAGCAGTAGCGACT 600
Db |||
QY 7616 GACTCTCATCGTATGAGCAACGCTTTTAAACAGGAGTTCATATAAGATTTTAAAGCAGTAGCGACT 600
Db |||
QY 601 STAGTTCTTCCAAAGTAAATCTTTGAGAGAAATTTTACAGAGTATTTACAGATCATATGAG 660
Db |||
QY 7676 GTGTAATTCCTAGCGGTCTCTACGCGAATTTTACGCGTATTTACAGATCATATGAG 660
Db |||
QY 661 ACCTTGAGGTATTTTCTCAACAGCCAAATCTTTGAGAGTATTTTACAGAGTATTTTACAGATCATATGAG 720
Db |||
QY 7736 ACTGTAGAGATTTTCTTTCGCAATTAACCAATCTCTTTGAGAGGCAAAATATAGCTTC 7795
Db |||
QY 721 TATACAGCGCTCTTGAAGGAAATTTTCCGATACAGACCTTTTATTAATGACAGAAATTT 780
Db |||
QY 7796 TATACAGCGCTCTTGAAGGAAATTTTCCGATACAGACCTTTTATTAATGACAGAAATTT 780
Db |||
QY 781 GAGAGGAGGTGTTTCTCAATACCAATCTCTTCCGAGTATGGAAGTCCCTTCTTG 840
Db |||
QY 7856 AACACTACTATTTACTTTTAACTGTAAGTAACTTACGCGAGTCAATGAGGAGGCGCTCTT 7915
Db |||
QY 841 ATTTCTAATGCTATCAAAATCGTCTGTAAGTAACTTACGCGAGTCAATGAGGAGGCGCTCTT 7915
Db |||
QY 7916 TTAATCAAGTGCAGTCAAAATCGTCTGTAAGTAACTTACGCGAGTCAATGAGGAGGCGCTCTT 7915
Db |||
QY 901 GCTCATGTTAACTCACTGAGTGTGTAAGTAACTTACGCGAGTCAATGAGGAGGCGCTCTT 960
Db |||
QY 7976 GCCCATGTTCACTCTCCAGAGTGTGTAAGTAACTTACGCGAGTCAATGAGGAGGCGCTCTT 8035
Db |||
QY 961 TCTGTAGTATTAATCACTCACTGAGTGTGTAAGTAACTTACGCGAGTCAATGAGGAGGCGCTCTT 1020
Db |||
QY 8036 ACTGTGAGGATTTGACCAATGAGTGTGTAAGTAACTTACGCGAGTCAATGAGGAGGCGCTCTT 8095
Db |||
QY 1021 ATTTAAAGTGAACAGTAAATTTCAATTTCTTATCACCAGTTCGACCAATTCACCTTAACA 1080
Db |||
QY 8096 TTAATAGCGAAAGGTGACTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTAAGT 8155
Db |||
QY 1081 CCAGCGATGAGGAAAGTGTGTAAGTAACTTACGCGAGTCAATGAGGAGGCGCTCTT 1133
Db |||
QY 8156 CCAGCGATGAGGAAAGTGTGTAAGTAACTTACGCGAGTCAATGAGGAGGCGCTCTT 8208
Db |||
```

## RESULT 4

```
US-10-158-844-20
; Sequence 20, Application US/10158844
; Publication No. US20040029118A1
; GENERAL INFORMATION:
; APPLICANT: Kunsch et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-R
COMPUTER: Dell Latitude Pentium 3
OPERATING SYSTEM: Windows 98
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10158,844
FILING DATE: 03-Jun-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/961,527
FILING DATE: 1997-10-30
APPLICATION NUMBER: US 60/029,960
FILING DATE: 1996-10-31
ATTORNEY/AGENT INFORMATION:
NAME: Hyman, Mark J.
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB340P1D1
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 2138 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-10-158-844-20
```

```
Query Match 53.6%; Score 608.2; DB 16; Length 21338;
Best Local Similarity 71.1%; Pred. No. 2.9e-116;
Matches 805; Conservative 0; Mismatches 328; Indels 0; Gaps 0;

QY 1 ATGATTCAATTTTCAATTAATCGCACATTTATTTTTCATGCTTTAAATACAACTAAAGCT 60
Db |||
QY 7076 ATGATTCAATTTTCAATTAATCGCACATTTATTTTTCATGCTTTAAATACAACTAAAGCT 7135
Db |||
QY 61 GCTATTAGCACTAAAAATGCCATTCCTATTCTTTTCAATCAATAAAAAATGAAGTCACTTCT 120
Db |||
QY 7136 GCTATTAGTTCATAAATGCCATTCCTATTCTTTTCAATCAAGATTAATAAATGACGTCACCAAT 7195
Db |||
QY 121 ACAGGAGTAACTTTAAACAGGCTTAACGCTCAATATCAATTAATTAATTAATTAATTAATTAAT 180
Db |||
QY 7196 GAAGGTTATCTTTAAATGCTTCAATTTGCTTCAATTTGCTTCAATTTGCTTCAATTTGCTTCAAT 7255
Db |||
QY 181 AGTAATGAAATGCTGTTTAAACAGTGTGTAATCACTCTCCAGGAGTATTTTATTAAGAGTGTGTT 240
Db |||
QY 7256 AAAAATGAAGTCTGTTTGTGTAATTAATTTTCTTTAGTTCGATCTCTTTGAGCTTCT 7315
Db |||
QY 241 TTTTATTAATTAATTTTCAAGTTTCCAGATATTAGTATAAATGTTAAAGAAATTTGAA 300
Db |||
QY 7316 TTTTATTAATTAATTTTCAAGTTTCCAGATATTAGTATAAATGTTAAAGAAATTTGAA 7375
Db |||
QY 301 CAACACCAAGTGTGTTTAAACAGTGGTAATCAGAGATTACCTTTAAAGGAAAGATGTT 360
Db |||
QY 7376 CAAAATCAAAATGTTTAAACAGTGGCAATCAGAAATACCTTAAAGGAAAGATAGC 7435
Db |||
QY 361 GACAGTATCTCGTCTACAAGAGTATCAACAGAAATCCTTTGATTTTAAACAAAAA 420
Db |||
QY 7436 GAACAAATATCCAGAAATCCAAGAAATTTTACAGACGACTCTTTTAATACCTTGAACAAA 7495
Db |||
QY 421 TTATTGAAGTCTATTATGCTGAACAGCTTTTGCAGCCAGTTTACAAGAAAGTCTCT 480
Db |||
QY 7496 TTACTCAGAAATTAATGAACAGCTTTTGCAGTATACAAAGAGATCGTCCG 7555
Db |||
QY 481 ATTTTAAACAGGAGTTCATATTTGATTAAGTAATCATATAAGATTTTAAAGCAGTAGCGACT 540
Db |||
QY 7556 ATTTTAAACAGGAGTTCATATTTGATTAAGTAATCATATAAGATTTTAAAGCAGTAGCGACT 540
Db |||
QY 541 GACTCTCATCGTATGAGCAACGCTTTTAAACAGGAGTTCATATAAGATTTTAAAGCAGTAGCGACT 600
Db |||
QY 7616 GACTCTCATCGTATGAGCAACGCTTTTAAACAGGAGTTCATATAAGATTTTAAAGCAGTAGCGACT 600
Db |||
QY 601 GTAGTTCTTCCAAAGTAAATCTTTGAGAGAAATTTTACAGAGTATTTTACAGATCATATGAG 660
Db |||
```

7676	GT	CGTAATTCCTAGCGGTTCTCTACGGGAATTTTTCAGCGGTATTTACAGATGATATCGAA	77335
661	QY	ACCGTTGAGGTATTTTTCTCACCAAGCCAAATCTGTGTCAGAAGTGAACACATTTCTCTTTT	720
7736	Db	ACTGTAGAGATTTCTTTTGGCAATAACCAAAATCTCTTTTAGAAGCGAAAAATATTAGCTTC	7795
721	QY	TATACAGCGCTCTTAGAAGGAATAATCCCGATACAGACCGTTTATTAAATGACAGAAATTT	780
7796	Db	TATACTCGTCTCTAGAAGGAATACTCTCTGATACAGATCGCTTGATTCCAAAGAGACTTT	7855
781	QY	GAGACGGAGGTTGTTTTCAATATCCCAATCCCTTTCCGACAGCTATGGAACGTGCTCTTTTG	840
7856	Db	AACACTACTATTACTTTTAAATGTGGTAAACTTACGCCAGTCAATGGAGCGTCCCGTCTT	7915
841	QY	ATTCTTAATGCTACTCAAAATGGTACTGTGTTAAGCTTGAGATTACTCAAAATCATATTTCA	900
7916	Db	TTATCAAGTCGCACTCAAAATGGTACTGTGAAACTTGAATAAAGGATGGGTGTTGTAGC	7975
901	QY	GCTCATGTTAACTCACCTGAGGTTGGTAAAGTAAAAGAGGATTTAGATATTTGTTAGTCAG	960
7976	Db	GCCCATGTTCACTCTCCAGAAGTTGGTAAAGTAAAAGAGAAATCGATCTGATCAGGTT	8035
961	QY	TCGTGGTAGTGAATTAATCTACGCTTCAATCCAACTTAATCTTAATGAGTCTTTTAAAGCT	1020
8036	Db	ACTGGTGAAGATTTGACCAATTAGTTTCAACCCCAACTTACTTGATTGATTTCTCTTAAAGCT	8095
1021	QY	ATTAAAAAGTCAAAACAGTAAAAAATCTTCTTATCTACAGGTCGACCAATTCACCCCTAAACA	1080
8096	Db	TTAAATAGCAAAAGGTGACTATTAGCTTTATCTCAGCTGTTGTCCTAATTTACTCTTGTCG	8155
1081	QY	CCAGGCGATGAGGAAGAAAGTTTTATCCAAATTAATACACCGGTACGAAACAA	1133
8156	Db	CCAGCAGATCTGACGAAGACTTCATGCAAGCTCAATCAACAGTTGCTGACAA	8208

## RESULT 5

```

US-09-815-242-9348
; Sequence 9348, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9348
; LENGTH: 1137
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae

```



Sequence 6810, Application US/09815242  
 Patent No. US20020061569A1  
 GENERAL INFORMATION:  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Kari L.  
 APPLICANT: Zyskind, Judith W.  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John D.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.  
 APPLICANT: Xu, H. Howard  
 TITLE OF INVENTION: Identification of Essential Genes in  
 FILE REFERENCE: ELITRA.011A  
 CURRENT APPLICATION NUMBER: US/09/815,242  
 CURRENT FILING DATE: 2001-03-21  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 NUMBER OF SEQ ID NOS: 14110  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 6810  
 LENGTH: 1131  
 TYPE: DNA  
 ORGANISM: Enterococcus faecalis  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (1)...(1131)  
 US-09-815-242-6810

Query Match 30.1%; Score 341.2; DB 9; Length 1131;  
 Best Local Similarity 57.0%; Pred. No. 4e-61;  
 Matches 644; Conservative 0; Mismatches 483; Indels 3; Gaps 1;  
 QY 4 ATTCAATTTTCAATTAATCGACATTATTTATTCATGCTTTTAAATACAACTAAAGTGCT 63  
 DB 1 ATGAATTTAATCGGTCAAGGAAGTCTTTTACGAAGATTACAACACTGTTCAACGACA 60  
 QY 64 ATTAGCACTAAAAATGCCATTCCTATCTTTCATCAATAAAAAATGAAGTCACTTCTACA 123  
 DB 61 ATTTCTTCTAAAAACAACGATTCCTTAAACAGGTGTTAAAAATGTTGTATCAGAAGAT 120  
 QY 124 GGAGTAATTTAAACAGGTCTAACCGTCAATATCAATTAATGAACACATTTCTCTAGT 183  
 DB 121 GCGTTATCACTTACTGTTGGAGTAAACCGGATATTTCAATTTGAAGTTTTTAAAGTAAGAT 180  
 QY 184 AATGAATGCTGGTTGCTAATACCTCTCCAGGAGCTATTTTATTAGAAGCTAGTTTT 243  
 DB 181 GATGAAGAAGCTCAATGACCATTTGAAACGACAGTAGCATCGTTTACATCTCGTTTC 240  
 QY 244 TTTATTAATATTATTCAAGTTTGCAGATATTAGTATAAATGTTTAAAGAAATTTGAACAA 303  
 DB 241 TTTCGGGAAATTTATTCGTAATATTACCAAGAAGATATGTTTCAATGGAAGTTCTAGATAAC 300  
 QY 304 CACCAAGTTGTTTAAACAGGTGTTAAATCAGAGATTACCTTAAAGGAAAGATGTTAC 363  
 DB 301 AATCAAGTTGCAATTTACTTTCAGGAAAGCTGATTTTACGGTTAATGGATTAGATCTGAC 360  
 QY 364 CAGTATCTCGTCTACAGAGATGATCAACAGAAAAATCTTTGATTTTAAAGAAATTTA 423  
 DB 361 AATTATCGCACTTACCAAGTAATGATATCTCAAAACCAATGAAATTTACCTGTTCAATTA 420

QY 424 TTGAAGTCTATTATGCTGAAACAGCTTTTTCGACGCCAGTTTCAAGAAAGTCGCTCTATT 483  
 DB 421 TTGACAAAAATATTAGTGAACAGGTTTCTGTATCGATCGACGAAAGTCGTCCTAATT 480  
 QY 484 TTAACAGAGTTTCAATATGTTAATTAAGTAATCATTAAGATTTTAAAGCAGTAGCGACTGAC 543  
 DB 481 TTAACAGAGTTTCAATATGTTAATTAAGTAATCATTAAGATTTTAAAGCAGTAGCGACTGAC 537  
 QY 544 TCTCATGCTATGAGCCAAACGTTTAAATCACTTTTGGACAAATCTTTCAGCAGATTTTATGTA 603  
 DB 538 TCACATCGTTTAAAGTCAACGTTGTATCCGACAGAAACAGCAGTAGAAGACTTTAAACATT 597  
 QY 604 GTTCTTCAAGTAATCTTTGAGAGAAATTTTTCAGCAGATTTTTCAGAGATGATTTGAGACC 663  
 DB 598 GTAATTCAGGAAAAAGTTTAACTGAACTTTCTCGTTTCAATTAACCAATGAAGAAAGATG 657  
 QY 664 GTTGAAGTATTTTCTCACCAGGCAAAATCTTGTTCAGAAAGTGAACATTTCTTTTAT 723  
 DB 658 GTTGAAGTATTTTCTCACCAGGCAAAATCTTGTTCAGAAAGTGAACATTTCTTTTAT 717  
 QY 724 ACAGCGCTCTTTAGAGGAAATTTATCCGATACAGACCGTTTATTAATGACAGAAATTTGAG 783  
 DB 718 TCTGTTTGTAGAGGAAATTTATCCTGATACCAACCGTTTAAATCCAAGTACCCATAAC 777  
 QY 784 ACGAGGTTGTTTCAATACCCAAATCCCTTCGCCACGCTATGGAACGTCCTTTCTGATT 843  
 DB 778 ACGCAATTTGAATTTTATGTACCAAGATTTGTTTTCAGCAATCGAACGTCCTTTTACTT 837  
 QY 844 TCTAATGCTACTCAAAATGTTGTTTAAAGTAAAGGATTAACGAGGATTTAGATATTGTTAGT 903  
 DB 838 TCTCATGAAGGACGTAAATATTTGTTGCGCTTTCAATTTCAACAGATTTCTGTTGTTTAA 897  
 QY 904 CATGTTAACTCACCTGAGGTTGGTAAAGTAAAGGATTAACGAGGATTTAGATATTGTTAGT 963  
 DB 898 TATGGAATTTCACTGAAATTTGGAAGGCTTGAAGAGCTTTTAACTATGAAAAATGTTTCT 957  
 QY 964 GGTAGTGAATTAATCACTATCAGCTTCAATCCAACTTACCTTATGAGTCTTTTAAAGCTATT 1023  
 DB 958 GGTGAAGCGTTGGATATTTCTTTCAACCCAGATATATGAAAGATGCGTTGCGAGCGTTT 1017  
 QY 1024 AAAAGTGAACAGTAAAAATTTCAATTTTATCACCAGTTTCAGCCATTTCACCTTAACACA 1083  
 DB 1018 GGCATATGAATATTTACCGTGAATTTCTTCAATTTCTGTTTACATTTGAGGACA 1077  
 QY 1084 GGCATAGGAGGAAAGTTTATCCAAATTAATTAACCCAGTACGAAACAAA 1133  
 DB 1078 ACCGAACAGAACTAGATTTCAATTTCACTAATTTACACCGGTTCTGTACAAA 1127

RESULT 8  
 US-10-282-122A-20800  
 Sequence 20800, Application US/10282122A  
 Publication No. US20040029129A1  
 GENERAL INFORMATION:  
 APPLICANT: Wang, Liangsu  
 APPLICANT: Zamudio, Carlos  
 APPLICANT: Malone, Cheryl  
 APPLICANT: Haselbeck, Robert  
 APPLICANT: Ohlsen, Kari  
 APPLICANT: Zyskind, Judith  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John  
 APPLICANT: Carr, Grant  
 APPLICANT: Yamamoto, Robert  
 APPLICANT: Forsyth, R.  
 APPLICANT: Xu, H.  
 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 FILE REFERENCE: ELITRA.034A  
 CURRENT APPLICATION NUMBER: US/10/282,122A  
 CURRENT FILING DATE: 2003-02-20  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848



; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 20800  
 ; LENGTH: 1128  
 ; TYPE: DNA  
 ; ORGANISM: Enterococcus faecalis  
 US-10-282-122A-20800

Query Match 29.8%; Score 338; DB 16; Length 1128;  
 Best Local Similarity 56.8%; Pred. No. 1.8e-60;  
 Matches 642; Conservative 0; Mismatches 485; Indels 3; Gaps 1;

QY 4 ATTCAATTTTCAATTAATCGCATTTATTTATTCATGCTTTTAAATACAACTAAACGTGCT 63  
 DB 1 ATGAAATTAAACGGTCAACGAAGTGTCTTTTACGAAGATTACAACTGTTTCAACGAGCA 60  
 QY 64 ATTAGCACTAAAATGCCATCTCTATTTCTTCATCAATAAAAAATGAAAGTCACTTCTACA 123  
 DB 61 ATTCTCTTAAACCAAGATTCCAATCTTAACAGGTGTAAATTTGTGTATCAGAAGAT 120  
 QY 124 GAGATTAATTTAACHGGGTCTAACGGTCAATATCAATATGAAACACATTTCTGTAAAT 183  
 DB 121 GGTCTATCACTTACTGTTGGGATTAACGGGATATTTCAATTTGAAAGTGTTTTAAAGTAAAGAT 180  
 QY 184 AATGAAATGCTGGTGTGCTAATTAACCTCCAGAGCTATTTTATTAAGAAGCTAGTTT 243  
 DB 181 GATGAAAGCTCAATGACCATTTGAACGCAAGGTAGCATGTTTACAACTCTGTTTC 240  
 QY 244 TTTATTAATATTTTCAAGTTTGCAGATATAGTATATAATGTTTAAAGAAATTTGAACAA 303  
 DB 241 TTTGCGGAAATTTATTCGTAATTTACGAAGATATGTTTCAATGGAAGTCTAGATAAC 300  
 QY 304 CACCAAGTTGTTTAAACCAAGTGTAAATCAAGATTTACCTTAAAGGAAAGATGTTGAC 363  
 DB 301 AATCAAGTTGCAATTTACTTCAAGAAAGCTGATTTTACGGTTAATGAGTATGCTGAT 360  
 QY 364 CAGTATCTCGTCTCAAGAAGTATCAACAGAAATCTTTGATTTTAAACCAAAATTA 423  
 DB 361 AATATCCGCACTTACCAAGTATGATATCTCAAAACCAATGAAATTAACCTGTTCAATTA 420  
 QY 424 TTGAAGTCTATTTGCTGAACAGCTTTTCAGCCAGTTTTACAGAAAGTCGTCCTATT 483  
 DB 421 TTGCAAAATTTATGATGAACAGGTTTCTGCTGATGATGATGATGATGATGATGATGAT 480  
 QY 484 TTAACAGAGTTCATATTTGATTAAGTAAATCATATAAGATTTTAAAGCAGTAGCGACTGAC 543  
 DB 481 TTAAGTGGGTTCACTTTATTTAGAAATCAAAA--TTACTTGGCTTGGCAGAGAT 537  
 QY 544 TCTATCGTATGAGCCAAAGTTTAACTCACTTTGACAAATATCTCAGCAGATTTGATGTA 603  
 DB 538 TCACATCGTTTAAAGTCAACGTGTGATCCGACAGAAACAGCAGTAGAAGCTTTAACATT 597  
 QY 604 GTTCTTCAAGTAAATCTTTGAGAGATTTTTCAGCAGTATTTTACAGATGATTTGAGACC 663  
 DB 598 GTAATTCAGGAAAGTTTAACTTAACTTCTCGTTCAATTAACCAATGAAGAAGAAATG 657

QY 664 GTTAGGATTTTCTCACCAGCCAAATCTGTTTCAAGAGTGAACACATTTCTTTTAT 723  
 DB 658 GTTGAATCAGCATTTATGGAATAATCAAGTGCTATTTAAACAGAGAAACAATGTACTTCTAT 717  
 QY 724 ACACGCCCTCTTAGAAGGAAATTTATCCGATACAGACCGTTTATTAATGACAGAATTTGAG 783  
 DB 718 TCTCGTTTGTAGAAGGAAATTTATCTGATACCAACCGTTTAAATTTCCAACTAGCCATAAC 777  
 QY 784 ACGGAGGTGTTTTCATATPACCAATCCCTTCGCCACGCTATGGAACGTCCTTCTTGTATT 843  
 DB 778 ACGCAATTTGAATTTTATGTACCAGAAATTTCTTTCAGCAATCGAACGTCCTTCTTACTT 837  
 QY 844 TCTAATGCTACTCAAAATGGTACTGTTTAAGCTTGAGATTAATCAAAATCATATTTACGCT 903  
 DB 838 TCTCATGAAGACGTAACAATATTTGTCGCTTTCAATTTCCACGAGATTTCTGTGTTTGA 897  
 QY 904 CATGTTAACTCACCTGAGGTTGTAAGTAAACAGAGATTTAGATATTTGTTAGTCAGTCT 963  
 DB 898 TATGMAATTCACCTGAAATTTGGAAGTGAAGAGCTTTAACTATGAATAATGTTTCT 957  
 QY 964 GGTAGTGAATTAATCACTAGCTTCAATCCAACTTACCTTATTTAGTCTTTTAAAAAGCTATT 1023  
 DB 958 GGTGAAGCGTTGGATATTTCTTTCACCCAGATTTATATGAAGATGCGTTGCGAGCGTTT 1017  
 QY 1024 AAAAGTGAACAGTAAATAATTTCAATTTCTTATCACCAGTTCGACCATTACCCCTAACACCA 1083  
 DB 1018 GGCATATGATATTAACCGTGAATTTCTTCTCAATTCGTCGTTTACATTTGGAGCCA 1077  
 QY 1084 GGCATAGGAAGAAAGTTTATCCAAATTAATTAACACAGTACGACACAAA 1133  
 DB 1078 ACCGAAACAGAACTAGATTTTCAATTAATTAACACCGGTTCTGTACAAA 1127

## RESULT 9

US-09-070-927A-622  
 ; Sequence 622, Application US/09070927A

; Patent No. US20020120116A1  
 ; GENERAL INFORMATION:

; APPLICANT: Charles A. Kunsch  
 ; Steven Barash  
 ; Patrick J. Dillon

; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides  
 ; NUMBER OF SEQUENCES: 982

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

; COMPUTER: HP Vectra 486/33

; OPERATING SYSTEM: MSDOS version 6.2

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/070,927A

; FILING DATE: 04-May-2000

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/046,655

; FILING DATE: 1997-05-16

; APPLICATION NUMBER: 60/044,031

; FILING DATE: 1997-05-06

; APPLICATION NUMBER: 60/066,009

; FILING DATE: 1997-11-14

; ATTORNEY/AGENT INFORMATION:

; NAME: Kenley K. Hoover

; REGISTRATION NUMBER: 40,302

; REFERENCE/DOCKET NUMBER: PB369

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 622:

SEQUENCE CHARACTERISTICS:

LENGTH: 8001 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 622:

US-09-070-927A-622

Query Match 29.2%; Score 330.6; DB 9; Length 8001;  
Best Local Similarity 56.9%; Pred. No. 1.5e-58;  
Matches 645; Conservative 0; Mismatches 485; Indels 4; Gaps 2;

QY 1 ATGATTCAATTTTCAATTAATCGCACATTAATTAATTCATGCTTTTAAATACAACTAAACGT 60  
DB ATTATGAAATTAACGGTCNAACGAAGTGTCTTTTACAAAGAATTACAAACTGTTTCAACGA 657

QY 61 GCTATTAGCACTAAAAATGCCATTCCTATTCTTT-CATCAATAAAAAATGAAGTCACCTTC 119  
DB GCAATTTCTTCTAAACCAACGATTCCTCAATCTTAACAGGTGTAAATTTGTTTATCAGA 717

QY 120 TACAGGAGTAACCTTTAACAGGGTCTAACGGTCAATATCAATTTGAAACACACTATTCTCGT 179  
DB AGATGGCTTATCACTTACTGAGGATACCGGATATTTCAATTTGAAGTTTTTAAAGTAA 777

QY 180 AAGTAATGAAATGCTGGTTTGCTAATTAACCTCTCCAGGAGCTATTTTATTAGAGCTAG 239  
DB AGATGATGAAAAAGCTCAATGACCATTTGAACGCACAGTAGCATCGTTTACAACTCTG 837

QY 240 TTTTATTATTAATTTATTTCAAGTTTGCAGATATTAGTATAATGTTAAAGAAATTGA 299  
DB TTTCTTTCGGCGAAATTTATTCGTTAAATTTACCAAGATATGTTTCAAAATGGAAGTTCTAGA 897

QY 300 ACAACACCAAGTTGTTTAAACAGTGGTAAATTCAGAGATTACCTTAAAGGAAAGAGTGT 359  
DB TAAACAACTCAAGTTGCAATTTACTTCAGAAAAGCTGATTTACGGTTAATGATAGATGC 957

QY 360 TGACAGATATCTCGTCTACAAGAAGTATCAACAGAAAATCTTTGATTTTAAAAACAAA 419  
DB TGATAATTTATCGCACTTACCAGTAATGATCTCAAAACCAAAATGAAATTTACCTGTTC 1017

QY 420 ATTATTGAAGTCTATTATTTGCTGAAACAGCTTTTCAGCCAGTTTACAAGAAAGTCGTC 479  
DB TTTATTGCAAAAATTTATTAGTGAACACAGGTTTTCGTATCGATGCACGAAAGTCGTC 1077

QY 480 TATTTTAAACAGGAGTTTCAATTTGTTAATGTAATCATAAAGATTTTAAAGCAGTAGCGAC 539  
DB AATTTTAACTGGGTTCACTTTATTTTAGAAAATCAAAA---TTACTTGGCGTTGCGAC 1134

QY 540 TGAATCTCATGATGAGCCAACTTTTAACTTTGACATATCTTGGACATATCTTACGAGATTGAT 599  
DB AGATTCACATCGTTTAAAGTCAACGGTGTATCCCGACAGAAACAGCAGTAGAAGACTTTAA 1194

QY 600 GGTAGTCTTCAAGTAATCTTTGAGAGAAATTTTCAGCAGTATTTCAGATGATATTGA 659  
DB CATTTGTAATTCAGGAAAAAGTTTAATGAACTTTCTCGTTCAITTAACCAATGAAGAAGA 1254

QY 660 GACCGTTCAGGTTATTTTCTCACCAAGCCAAATCTTTGTCAGAAGTGAACATTTCTTT 719  
DB AATGGTTGAAATCAGCAATATGGAATAATCAGTGCTATTTAAACAGAAACAAATGACTT 1314

QY 720 TTATACAGCCTCTTTAGAGGAAATTTATCCCGATACAGACCGTTTATTAATGACAGAAAT 779  
DB CTATTTCTGTTTGTAGAGGAAATTTATCTGATACCAACCGTTTAAATTCGAACCTAGCA 1374

QY 780 TGAGACGGAGGTTGTTTTCATACCAATCCCTTCGCCACGCTATGGAACGTCCTTTCTT 839  
DB TTAACACGCAAAATGAAATTTTATGTATCCAGAAATTTGCTTTTTCAGCAATTCGAACGTCCTTT 1434

QY 840 GATTTCTTAATGCTACTCAAAATGGTACTGTTTAAAGCTTGAGATTACTCAAAATCATATTTC 899  
DB ACTTCTCATGAAGGACGTAAACAATATTGTTTCGCTTTTCAATTTTACCAGATTTCTGTGT 1494

## RESULT 10

US-10-282-122A-21823

; Sequence 21823, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 21823

; LENGTH: 1128

; TYPE: DNA

; ORGANISM: Enterococcus faecium

US-10-282-122A-21823

Query Match 27.8%; Score 315.6; DB 16; Length 1128;

Best Local Similarity 55.6%; Pred. No. 8.7e-56;

Matches 628; Conservative 0; Mismatches 499; Indels 3; Gaps 1;

QY 4 ATTCATTTTCAATTAATCGCACATTTATTTATTCATGCTTTTAAATACAACTAAACGTGCT 63

```
Db 1 ATGAAAGTACTTTTAAACCGAGCTAGCTTTATGCGGAATTCGAACCTGTTCAACGAGCT 60
QY 64 ATTAGCACTAAAATGCAATTCCTATTCTTTCTCATCAATAAAATGGAAGTCACTTCTACA 123
Db 61 ATTTCAAGCAAAACCGAGTCCCTATTATTTGACAGGTGTAAATAATCACTGACACAAGAA 120
QY 124 GGAAGTAACTTTAAACAGGGTCTAAACGGTCAAAATATCAATTAAGAAACACTTCTCTGAAGT 183
Db 121 GGTGTTGACTTTGACGGGAGCAACGCTGATATATCAATTAAGAACTTTTGTCTGTGAA 180
QY 184 AATGAATACTGCTGTTGCTAAATTAACCTCCAGGAGCTATTTTATTAGAAGCTAGTTTTT 243
Db 181 AAGCAAAAAGCAAAATATGCAAAATCGAATCTACTGTTCCATTTGTTTACAAGCAGCTTC 240
QY 244 TTTATTAAATATTTTCAAGTTTGCCAGATATTAGTATTAATGTTTAAAGAAATGGAACA 303
Db 241 TTTAGCGAATCAATCGGAGACTTCTCGAAGAACATTTACTTTAGAACTTTTAGAAAT 300
QY 304 CACCAAGTTGTTTAAACCAAGTGTAAATCAGAGATTACCTTAAAGGAAAGATGTTGAC 363
Db 301 AAACAAGTAGCGATCACTCTCGAAAGCGAATTTATCGTAATGGATTAGATGAGAT 360
QY 364 CAGTATCCTCGTCTACAAGAGTATCAACAGAAATCTCTTGTATTTTAAATAAATAATTA 423
Db 361 AACTATCCTCATCTCTCTGTGTCGAAGGCCATAACAGATGAAATTAACCTGTACAGTA 420
QY 424 TTGAAGTCTATTATTCTGAAACAGCTTTTGCAGCCAGTTTACAGAAAGTGCCTTATT 483
Db 421 TTGACTAACTAATCAACGAAACAGTTTTTGTCTCTCAACATGAGAGTCTGTCATC 480
QY 484 TTAACAGGAGTCTATATGTTAAAGTAAATCAATAAGATTTTAAAGCAGTAGCGACTGAC 543
Db 481 TTGACAGGTGTGCAATTTCAATTTA---TCTGATAATCTTTTATTAGCTAGTACTGAT 537
QY 544 TCTCATGCTATGAGCAAGTGTATCACTTTGGACAATCTTCAGCAGATTTGATGTA 603
Db 538 TCTACCGTCTAAGTCAACGCGTGAATCCAGTAGAACAAAGCGGCTGATCATTTTGATAT 597
QY 604 GTTCTTCCAAGTAAATCTTTGAGAGAAATTTTCAAGAGATTTTACAGATGATATTGAGACC 663
Db 598 GTTATTCCTGGAAAAAGTTTGATCGAATTAATCTCGCTATTAACAAATGAAGAAGATC 657
QY 664 GTTGAGGATTTTCTCACAAGCAAAATCTTGTGTTGAGAAGTGAACACATTTCTTTTAT 723
Db 658 GTTCAAAATCAGCATTTATGAAACCAAGTGTGTTTCAAAACAGAAACGATGATTTCTAT 717
QY 724 ACACGCTCTTGAAGGAAATTTATCCGATACAGACCGTTTATTAATGACAGAAATTCAG 783
Db 718 TCCCGTTTGTAGAAGGAAATCTATCCAGATACCAATCGTTTGAATTCCTTCAAGTTTAA 777
QY 784 ACGGAGTGTGTTTCAATACCAATCCCTTCCGACGCTATGGAACGTCGCTTCTTGATT 843
Db 778 ACAGAGTGTGAATTTCTGTTCTAGCTTTTACGGGATCGAAGCGGCTCTTTACTT 837
QY 844 TCTAATGCTACTCAAAATGCTACTGTTAAGCTTGAGATCTCAAAATCATATTTACGCT 903
Db 838 TCTCATGAAGCGGTAACAACATCGTTCGTTTATCTATTCGTCAGATGCGGTTGTTTA 897
QY 904 CATGTTAACTCACCTGAGGTTGTTAAGGTAAACAGAGGATTTAGATATTGTTAGTCT 963
Db 898 TAGCGAAATTCGCCAGAAATCGGNAAGTAGAAGAAAGCTTGAGTTATACAGCAAGCAGT 957
QY 964 GGTAGTGATTTAACTATCAGCTTCAATCCAACTTACCTTATGAGTCTTTTAAAGCTATT 1023
Db 958 GCGGATCCATTAGATATTTCTTCAATCCGATTTATATGAAGCAGCGGTTGCGTGCATTT 1017
QY 1024 AAAAGTGAACAGTAAATAATTCATTTCTTATCACAGTTCGACCAATTCACCTTAACCA 1083
Db 1018 GGAGATGAGCAACAAGTGAATTTATCTCTGCTATTCGTCATTTACATTTAGGAACCA 1077
QY 1084 GCGGATGAGGAAGAAAGTTTATCCAAATTAATTAACCAAGTACCAACAAA 1133
```

```
Db 1078 ACGGAAGATGGCTTCAATTTATTTCAGCTGATTACTCTCTGTCGTCACAAA 1127

RESULT 11
US-10-282-122A-34649
; Sequence 34649, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34649
; LENGTH: 1131
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-10-282-122A-34649
```

```
Query Match 21.8%; Score 247.4; DB 16; Length 1131;
Best Local Similarity 52.9%; Pred. No. 1.4e-41;
Matches 580; Conservative 0; Mismatches 511; Indels 6; Gaps 2;

QY 1 ATGATTCATTTTCAATTAATCCACATTTATTTATTCATGCTTTAAATCAACATAACGT 60
Db 1 ATGATGGAATTCATTAATAAGAGATTTATTTATTAATCAACTTAACGACATTTAAAG 60
QY 61 GCTATTAGCACATAAAATGCCATTTCTTATTTTTCATCAATAAAATTCGAAGTCACTTCT 120
Db 61 GCTATCTCACCTAGAACAACTTTACCGATTTTAAACAGGTATCAAAATGATGCTAAAGAA 120
QY 121 ACAGGAGTAACTTAAACAGGCTTAACGGTCAAAATATCAATTAATCAATTAATTCCTTC 177
Db 121 AATGAAGTCAATTTTAACTGGGTGAGATTTCAGAAATATCAATAGAAATCACTATTCCAAA 180
QY 178 GTAAGTAATGAAATGCTGTTTGTCTAATTAATCTCTCCAGGAGCTATTTTATTAGAGCT 237
Db 181 CAAGTTGATGGTGAAGAAATTTGTAATTAACAGAAACAGGATCAGTTCTTCTCTGCT 240
```

```
QY 238 AGTTTTTTTAAATATTTTCAAGTTTGCAGATATAGTATATAATGTTAAAGAAAT 297
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 CGTTTCTTCGTTGATATTTAAATAAATCTCTCGAAAAGAGTTAAATATCAACTAAT 300
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 298 GAAACAACCAAGTGTGTTTAAACAGTGTGTAATCAGAGATTACCTTAAAGGAAAGAT 357
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 GAACAATTCACAAACGTTAATCACATCAGTGTATTCAGAAATTTAACTTAAGTGGCTTAGAT 360
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 358 GTTGACAGTATCTCGCTCAACAAGATATCAACAGAAAATCTCTTTGATTTTAAACA 417
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 CCTGATCAATACCCATTTATCTAGTGTATCAGAGATGACGCTATTCAATTTGCTAGT 420
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 418 AAATTTTGAAGTCTATTATCTGTAACAGCTTTTGCAGCAGTTTACAGAAAGTGT 477
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 AAGTACTAAAAAATATCATGTCACAAATTAATTTGCAAGTGTCCACCTCAGAAAACGA 480
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 478 CCTATTTTAAACAGGAGTTCATATCTGTTAAGTAATCATAAAGATTTTAAAGCAGTAGCG 537
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 481 CCAGTACTTACTGGTGTAACCTGGCTTATA---CAAGATATGAATTAATATGCACAGA 537
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 538 ACTGACTCTCATCTGATGAGCAACGTTTAAATCACTTTGGCAATATCTTTCAGCAGATTTG 597
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 538 ACAGATTCACACCGCTTAGCTGTAAGAAAGTTACAGTTAGAGATGAATCAGAAATATAA 597
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 598 ATGGTAGTCTTCCAAAGTAAATCTTTGAGAGAAATTTTCAGCAGTATTACAGATGATATT 657
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 598 AATGTCATCATCTCTGGTAAAGCTTTATCTGAAATTAACAAATATGAGTGACAGCAG 657
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 658 GAGACCGTTGAGGTATTTTCTCACCAAGCCAAATCTTGTTCAGAAAGTGAACACATTTCT 717
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 658 GAAGATATTTGATATTTCTTCTTCTTAACCAAGTGTATTTCAGAGTGGGNAATTAAT 717
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 718 TTTTATACAGCGCTCTTGAAGAAATATATCCCGATACAGACCGTTTATTAATGACAGAA 777
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 718 TTTATCTCACGTTTACTTGAAGGTCAATATCCAGATACGACACGTTTATCCAGAAAAT 777
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 778 TTTGAGCGGAGTGTGTTTCAATACCAATCCCTTCGCCAGCTATGGAAGTGCCTTC 837
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 778 TATGAGATTAATAGGAATTAACAAATGGAGATCTTATCATGCAATGATCGTGCATCT 837
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 838 TTGATTTCTAATGCTACTCAAAATGGTACTGTGTAAGCTTGAGATTACTCAAAATCATATT 897
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 838 TTATTAGACGTGAAGTGGAAATATGTTTATTAATTAATTAAGTACAGGTAATGAATTAGTT 897
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 898 TCAGTCAATGTAATCACTGAGGTGTTGTAAGGTAACGAGGATTTAGATATTGTTAGT 957
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 898 GAACTTTCTACTCTCTCTGAAATTTGTTACTGTGTAAGGAAAGTAAACGCTAATGAT 957
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 958 CAGTCTGCTAGTATTAATCAATCAGCTTCAATCCAACTTACTTATGAGTCTTTAAA 1017
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 958 GTAGAAGCGGAAACTTGAATAATTTCTTTCAACTCAAAATCATGATGATGCTTTAAA 1017
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1018 GCTATTAAGAGTGAACAGTAAATAATTCATTTCTTATCACCAGTTCGACCATTCACCTA 1077
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1018 GCCATTGATATGATGAAGTAGAGTAGATAGTATCTTTGTTGTAACATGAACCAATTTATCTTA 1077
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1078 ACACAGCGGATGAGA 1094
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1078 AAACCAAAAGATGATGA 1094
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

## RESULT 12

```
US-10-282-122A-35724
; Sequence 35724, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
```

```
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 35724
; LENGTH: 1131
; TYPE: DNA
; ORGANISM: Staphylococcus haemolyticus
US-10-282-122A-35724
```

```
Query Match 21.5%; Score 243.4; DB 16; Length 1131;
Best Local Similarity 52.8%; Pred. No. 9.9e-41;
Matches 598; Conservative 0; Mismatches 526; Indels 9; Gaps 3;
```

```
QY 1 ATGATTCNAATTTTCAATTAATCGCACATTTATTTATTCATGCTTTTAAATACAACTAAACGT 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1 ATGATGGAATTCACAAATTTAGAGAGATTTATTTTATTAATCAATTAATTAATGACACATTAATA 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 GCTATTAGCACTAAAAATGCCATTCCTTATCTTTCATCAATAAAAAATGGAAGTCACTTCT 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 GCCATCTCACAAGAACACACATTTACCAATTTTAAGGGGTATCAAAATCGATGCTTAAGAT 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 ACAGAGTAACTTTTAAACAGGCTTAACGGTCAAAATATCAATTTGAAAACACATTTCTCTGTA 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 AACGAAGTCACTTCTTACTGTTTCAAGATCTCTAGATATCTATTGAAATTTACAATCCCTAAA 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 AGTAATGAAATGCTGGTTTCTGCTAATTAACCTCT---CCAGGAGCTATTTTATTAAGAGCT 237
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 CAAGTAGATGGTGGAGATTTGTCATATTTCTGAAAACAGGTTCAAGTTGTTACTTCTCTGCT 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 238 AGTTTTTTTAAATATTTTCAAGTTTGCAGATATTAGTATATAATGTTTAAAGAAAT 297
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 CGTTTCTTCGATATATTTTAAATAAATCAACAGTAAAGATGTTAAATTTATCACAAT 300
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 298 GAACAACCAAGTGTGTTTAAACAGTGTGTAATCAGAGATTAACCTTAAAGGAAAGAT 357
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 GAACAATTTCAACACACTGATTACTTTCAGGACATTTCTGAATTTAACTTAAGTGGTTTAGAT 360
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 358 GTTGACAGTATCTCTGCTTACAGAAAGATCAACAGAAAATCTCTTTGATTTTAAACA 417
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 CCTGATCAATATCCATTTATTTACCTCAAGTATCAGCTGATGATGCAATTTCAATTTATCAGTA 420
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 418 AAATTTTGAAGTCTATTATTTGCTGAAACAGCTTTTTCAGCAGCTTTTACAGAAAGTGT 477
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 AAAGTATTAATAAATATATCAATTTGCAAAACAAATTTTCGCACTGTCACCTCAGAAAACGCG 480
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 478 CCTATTTTAAACAGGAGTTTATGTTAATTAAGTAATCATTAAGATTTTAAAGCAGTAGCG 537
Db 481 CCAGTACTTACTGGTGTAACTGGCTTATA--CAAGATAATGAATTAATATATCACTGCG 537
QY 538 ACTGACTCTCATCGTATGAGCCACGTTTAAATCACTTTGGACATACTTCAGCAGATTG 597
Db 538 ACCGATTACACCGCTTGGCTGTGAAGAAAGTTAAATTAGAAGATGACCTGAAACAAA 597
QY 598 ATGGTAGTCTTCCAGTAATCTTTGAGAGAAATTTTCAGCAGATTTTACAGATGATATT 657
Db 598 AATGTCATCATCTCCAGTAAAGCTTTATCAGAAATTAATAAATTAATGATGATGCGAT 657
QY 658 GAGACCGCTTGAGGTATTTTCTCAACGACCAATCTTGTTCAGAGTGAACACATTTCT 717
Db 658 GAAGAGATTGATATTTCTTCTGCTCAATCAAGTTTATTTAAAGTTGAAATGTAAC 717
QY 718 TTTTATACAGCTCTTAGAAGAAATTTATCCGATACAGACGTTTATTAATGACAGAA 777
Db 718 TTTTATTTCTGCTTATTAGAAGACATTTATCTGTATACACACGTTTATTCCTGAGAT 777
QY 778 TTTGAGACGAGGTTTCTCAATACCAATCCCTTCGCCACGCTATGGAACGTGCTTC 837
Db 778 TAGCAATATAATAGCTTAGACAACGGTGAATCTATCAGCAATCGATCGTCTCT 837
QY 838 TTGATTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTTGAGATTACTCAAAATCATATT 897
Db 838 TTATTTGCCAGGAGGTTGCAACATGTTATTAGTTAAGTACGGTAAATGATGTTGA 897
QY 898 TCAGCTCATGTTAACTCACTGAGGTGGTAAAGTAAACGAGATTTAGATATTTAGT 957
Db 898 GAATTAATCACTACGTCACCTGAGATTTGTTACTGTAAAGAAAGAAAGTTACAGCAACGAT 957
QY 958 CAGTCTGTGAGTGAATTAATCACTATCAGCTTCAATCAACTTACTTTATGAGTCTTAA 1017
Db 958 GTAGAGGTTGGCACTTGAAGATTCTTCAATCTCGTTATATGATGATGCGTTAA 1017
QY 1018 GCTATTAAGTGAACAGTAAATTTCAATTTCTTATCACCAGTTTCGACCAATTTACCCCTA 1077
Db 1018 GCGATTGATAATCATGAGTGAAGTTGAATTTCTTTGGTACGATGAACCGTTTATCCT 1077
QY 1078 ACACAGGCGGATGAGGAGAAAGTTTATCCCAATTAATTAACACGATGAGAAC 1130
Db 1078 AARCCA--AAAGAGATGATTTCTGTCTCAATTAATTTTACCCAATTAGAAC 1127
```

```
RESULT 13
US-09-815-242-8628
; Sequence 8628, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Walli, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815.242
; PRIOR FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
```

```
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8628
; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1134)
; US-09-815-242-8628

Query Match      20.8%; Score 236.2; DB 9; Length 1134;
Best Local Similarity 52.2%; Pred. No. 3.1e-39;
Matches 573; Conservative 0; Mismatches 518; Indels 6; Gaps 2;

QY 1 ATGATTTCAATTTTCAATTAATCGCACATTTATTTATTCATGCTTTTAAATACAACTAAACGT 60
Db 1 ATGATGGAATTCACCTATTAAAGAGATTTATTTATTACACAATTAATGACACATTAA 60
QY 61 GCTATTAGCACTAAATGCCATTCCTATTCTTTTCATCAATATAAAATTTGAAGTCACATTCT 120
Db 61 GCTATTTTACCACCAAGACAACATTTACTTATTAATCTGTTATCAAAATTCGATGCGAAAGAA 120
QY 121 ACAGAGTAACCTTTAAACAGGGTCTAACGGTCAATATCAATTTGAACACATTTTCTCTGTA 180
Db 121 CATGAAGTTATTTAACTGGTTCAGACTCTGAAATTTCAATAGAAATCACTATTCTCTAAA 180
QY 181 AGT---AATGAATGCTGGTTGCTAATTAACCTCTCCAGGAGCTATTTTATTAGAAGCT 237
Db 181 ACTGTAGTACGGAAGATATTGTCAATATTTTCAGAAACAGGCTCAGTAGTACTTCTCTGGA 240
QY 238 AGTTTTTTTATTAATTTTCAAGTTTTCGCGAGATTTAGTATATAATGTTAAAGAAATTT 297
Db 241 CGATTTCTTTGTTGATTTATAAAATTAACCTGGTAAAGATGTTAAATTTATCTACAAAT 300
QY 298 GAACAACACCAAGTTGTTTAAACCAGTGGTAAATTCAGAGATTACCTTTTAAAGGAAAGAT 357
Db 301 GAACAATTTCCAGACATTAATTTACATCAGGTCATTCTGAATTTAATTTAGTGGCTTAGAT 360
QY 358 GTTGACCACTATCTCGCTACAGAGATTAACAGAAATTCCTTTGATTTTAAACAA 417
Db 361 CCAGATCAATATCTTTTATTTACCTCAAGTTTCTAGAGATGACGCAATTTCAATTTGTCGTA 420
QY 418 AAATTTATGAAGTCTATTATTGCTGAAACAGCTTTTTCAGCCAGTTTACAGAAAGTCGT 477
Db 421 AAAGTCTTAAACACGTTGTCACAAACAAATTTTTCAGTGTCCACCTCAGAAACACGC 480
QY 478 CCTATTTTAAACAGGATTCATATTGTTAAGTAAATCATTAAGATTTTAAAGCAGTAGCG 537
Db 481 CCAGTACTAATCTGTTGAACTGGCTTATA---CAAGAAATTAATTAATATGACACAGCG 537
QY 538 ACTGACTCTCATCGTATGAGCCACGTTTAAATCACTTTTCGACATACTTTACAGCAGATTG 597
Db 538 ACTGACTCACACCGCTTGGCTGTGAAGAAAGTTGCGATGAGAGATGTTTCTGAAACAAA 597
QY 598 ATGGTAGTCTTCCAAAGTAAATCTTTGAGAGAAATTTTCAGCAGATTTTACAGATGATATT 657
Db 598 AATGTCATCATTTCCAGTAAAGCTTTTAGCTGAATTAATAAATTAATGCTGACATGAA 657
QY 658 GAGACCGTTGAGGTATTTTCTCACCAACGCAATCTTCTCAGAGATGTAACACATTCTTCT 717
Db 658 GAAGCAATTTGATATCTTCTTCTGCTTCAACCAAGTTTATTTTAAAGTTGGAATGTAAC 717
QY 718 TTTTATACACGCTCTTTAGAGAAATTTATCCGATACAGACGCTTTTATTAATGACAGAA 777
Db 718 TTTTATTTCTCGATTATTAGAGGACATTTATCTGTATACACACGTTTATTCCTGNAAC 777
QY 778 TTTGAGACGAGGTTGTTTTCATATCCCAATCCCTTTCGCCACGCTATGGAACGTGCTTC 837
```

```

Db 778 TATGAAATTAATTAAGTATACAAATGGGAGTTTATCATGCGATTGATCGTGCCTCT 837
Qy 838 TTGATTTCTAATGCTACCAAAATGGTACTGTTAAGCTTTGAGATTACTCAAAATCATATT 897
Db 838 TTATTAGCGGTGAAGTGGTAAATACCTTATTAATTAAGTACAGGTGATGCGTGT 897
Qy 898 TCAGCTCATGTTAACTCACTGAGGTGGTAAAGTAAACGAGGATTAAGATATTGTTAGT 957
Db 898 GAAATGTCTTCTACATCAACCAAAATGGTACTGTAAAGAAAGAGTGTGATGCAACGAT 957
Qy 958 CAGTCTGCTAGTGAATTTAACTATCATGCTTCAATCAACTTACCTTATTGAGTCTTTAAA 1017
Db 958 GTTGAGGTGGTAGCGCTGAAATTTCAATCACTCTAAATATATGATGATGCTTTAAAA 1017
Qy 1018 GCTATTAAAGTGAACACAGTAAATTCATTTCTTATCACCAGTTTCGACCACTTACCCCTA 1077
Db 1018 GCAATCGATATGATGAGGTGGAAGTTGAAATTCCTCGGTACAAATGAAACCAATTTTCTA 1077
Qy 1078 ACACGAGCGGATGAGGA 1094
Db 1078 AAACCAAAAGGTGACGA 1094

```

RESULT 14

```

US-10-449-830A-1
; Sequence 1, Application US/10449830A
; Publication No. US20040137516A1
; GENERAL INFORMATION:
; APPLICANT: Phagotech Inc.
; APPLICANT: Dubow, Michael
; APPLICANT: Gros, Philippe
; APPLICANT: Bergeron, Dominique
; TITLE OF INVENTION: COMPOSITIONS AND METHODS INVOLVING AN ESSENTIAL STAPHYLOCOCCUS AU
; FILE REFERENCE: Q79017
; CURRENT APPLICATION NUMBER: US/10/449,830A
; CURRENT FILING DATE: 2003-05-31
; PRIOR APPLICATION NUMBER: PCT/CA01/01754
; PRIOR FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US 09/727,892
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 09/885,561
; PRIOR FILING DATE: 2001-06-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1134
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1134)
; OTHER INFORMATION:

```

US-10-449-830A-1

```

Query Match 20.8%; Score 236.2; DB 17; Length 1134;
Best Local Similarity 52.2%; Pred. No. 3.1e-39;
Matches 573; Conservative 0; Mismatches 518; Indels 6; Gaps 2;

Qy 1 ATGATTCAATTTTCAATTAATGCGACATATTATTATTCATGCTTTTAAATCAACATAACGT 60
Db 1 ATGATGGAATTCACATATTAAAAGAGATATTATTATACACAATTAATGACACATTAAAA 60

Qy 61 GCTATTAGCACTAAAATGCCATTCCTATTCTTTCATCAATAAATAAATGAGTCACTTCT 120
Db 61 GCTATTTCACCAAGAACACATTAATCACTATTAATGATGATCAAAATCGATGCGAAGAA 120

Qy 121 ACAGAGTAACCTTTAAACAGGGTCAACCGTCAAAATCAATTAATGAAAAACATTTCTCTGA 180
Db 121 CATGAAGTATTAATTAAGTGTTCAGACTCTGAAATTTCAATAGAAATCACTATTCTTAA 180

```

```

Qy 181 AGT---AATGAAATGCTGTTTCTTAATTAACCTCTCCAGGAGCTATTTTATTAGAACCT 237
Db 181 ACTGTAGTGGCGAAGATATTGTCAATATTTAGAAACAGGCTCAGTAGTACTTCTTGGGA 240
Qy 238 AGTTTTTTTATTAATAATTATTTCAGAGTTTGGCAGATATTAGTATAAATGTTTAAAGAAAT 297
Db 241 CGATTTCTTTGATATTTATAAAAAAATTAACCTGGTAAAGATGTTAAATTTATCTACAAT 300
Qy 298 GAACAACACCAAGTTGTTTTAACCAAGTGGTAAATCAGAGATTACCTTTAAAGGAAAGAT 357
Db 301 GAACAAATCCAGACATTAATTAACATCAGGTCAATCTGAATTTAAATTTAAGTGGCTTAGAT 360
Qy 358 GTTGACCAAGTATCTCTGTCTACAAGAAGTATACACAGAAATCTTTTGAATTTTAAAAACA 417
Db 361 CCAGATCAATATCTCTTTATTAACCTCAAGTTTCTAGAGATGACGCAATTTCAATTTGCGTA 420
Qy 418 AAAATTTTGAAGTCTATTATTGCTGAAACAGCTTTTGCAGCCAGTTTACAAAGAAAGTCT 477
Db 421 AAAGTCTTAAACAGGTGATGCAACAAATTTTGCAGTGTCCACCTCAGAAACACGC 480
Qy 478 CCTATTTTAAACAGAGTTTCATATTGTATTAAGTAAATCAATAAAGATTTTAAAGCAGTAGCG 537
Db 481 CCAGTACTAACTGGTGTGNACTGGCTTATA---CAAGAAATGAATTAATATGCACAGCG 537
Qy 538 ACTGACTCTCATGTATGAGCCCAAGTTTAACTACTTTTGGACAATCTTTCAGCAGATTGG 597
Db 538 ACTGACTCACACCGCTTTGGCTGTAAAGAAAGTTGCAAGTTAGAGATGTTTCTTGAACAAA 597
Qy 598 ATGGTAGTTCTTCAAGTAAATCTTTGAGAGAAATTTTTCAGCAGTATTTACAGATGATATT 657
Db 598 AATGTCAATCAATCCAGGTAGGCTTTAGTGAATTAATAAATTAATGCTCGACANTGAA 657
Qy 658 GAGACCGTTGAGGTATTTTCTCAAGCAAGCAAAATCTTTGTTTCAAGAGTGAACACATTTCT 717
Db 658 GAAGACATTTGATACTTCTTCTTCTTCAAAACCAAGTTTATTATAAGTTTGAATGTGAAC 717
Qy 718 TTTTATACACGCTCTTAGAAGGAAATTTATCCGATACAGACCGTTTATTATTAATGACAGAA 777
Db 718 TTTTATTTCTCGATTATTAGAGAGCAATTTATCTGATACACACGTTTATTTCCCTGANAAC 777
Qy 778 TTTGAGAGCGGAGGTGTTTTCATACCAATCCCTTCCGACGCTTATGGAACGTCCTTC 837
Db 778 TATGAATTAATTAAGTATAGACANTGGGAGTTTATCATGCGATTGATGCGTCTCT 837
Qy 838 TTGATTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTTGAGATTACTCAAAATCATATT 897
Db 838 TTATTAGCGGTGAAGGTGGTAAACGTTATTAAATTTAAGTACAGGTGATGACGCTGTT 897
Qy 898 TCAGCTCATGTTAACTCACTGAGGTGGTAAAGTAAACGAGGATTTAGATATTGTTAGT 957
Db 898 GAATTTGCTTCTACATCACCAAGAAATTTGGTACTGTAAAGAAAGAAAGTTGATGCAACGAT 957
Qy 958 CAGTCTGCTAGTGAATTTAACTATCATGCTTCAATCAACTTACCTTATTGAGTCTTTAAAA 1017
Db 958 GTTGAGGTGGTAGCGCTGAAATTTTCAATCACTTAAATATATGATGATGCTTTTAAA 1017
Qy 1018 GCTATTAAAGTGAACACAGTAAATTCATTTCTTATCACCAGTTTCGACCAATTTACCCCTA 1077
Db 1018 GCAATCGATATGATGAGGTGGAAGTTGAAATTTCTCGGTACAAATGAAACCAATTTTCTA 1077
Qy 1078 ACACGAGCGGATGAGGA 1094
Db 1078 AAACCAAAAGGTGACGA 1094

```

RESULT 15

```

US-08-781-986A-153
; Sequence 153, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255

```



```

CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
;
; INFORMATION FOR SEQ ID NO: 153:
; SEQUENCE CHARACTERISTICS:
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-08-781-986A-153

```

```

Query Match          20.8%; Score 236.2; DB 8; Length 2347;
Best Local Similarity 52.2%; Pred No. 4,3e-39;
Matches 573; Conservative 0; Mismatches 518; Indels 6; Gaps 2;

QY 1 ATGATTCATATTTCAATTAATCGACATATTATTTATTCATGCTTTTAAATACAACTAAACGT 60
DB 393 ATGATGGAATTCACATATTAAGAGAGATTTATTTATACAAATTAATGACACATTAATA 452
QY 61 GCTATTAGCATAAAATGCCATTCCTATTCCTTCATCAATAAATTAAGTCACTTCT 120
DB 453 GCTATTCACCAAGAACAACTATTAATTAATGCTGATCAAAATCGATGCGAAGAA 512
QY 121 ACAGAGTAACCTTTAACAGGGTCTAACGGTCAAATATCAATTTGAACACATATTCCTGTA 180
DB 513 CATGAAGTTATATTAACCTGGTTCAGACTCTGAATTTCAATAGAAATCACTATTCCTTAA 572
QY 181 AGT---AATGAAATGCTGGTTGCTAATTAATTAATTAATTAATTAATTAATTAATTAAT 237
DB 573 ACTGATAGATGGGAAGATTAATGCTCAATATTCAGAAACAGGCTCAGTAGTACTTCTCTGA 632
QY 238 AGTTTTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 297
DB 633 CGATTTCTTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 692
QY 298 GAACAAACCAAGTTGTTTTAACAGTGGTAAATCAGAGATTACCTTTAAAGAGGAAAGAT 357
DB 693 GAACAAATTCAGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 752
QY 358 GTTGACCAAGTATCCTCGCTACAGAGATTAATTAATTAATTAATTAATTAATTAATTAAT 417
DB 753 CCAGATCAATATCCCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 812
QY 418 AAATTTTGAAGTCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 477
DB 813 AAAGTGTCTTAAAGACGTTGATTCAGCAACAAATTTTGCAGTGTCCACTCAGAAACAGC 872
QY 478 CCTATTTTAAACAGAGTTCATATTTGATTAATTAATTAATTAATTAATTAATTAATTAAT 537
DB 873 CCAGTACTACTGGTGAACCTGGCTTATA---CAAGAAATTAATTAATTAATTAATTAATTAAT 929

```

```

RESULT 16
US-10-329-624-153
; Sequence 153, Application US/10329624
; Publication No. US20040043037A1
;
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
;             Gil H. Choi
;             Patrick S. Dillon
;             Craig A. Rosen
;             Steven C. Barash
;             Michael R. Pannon
;
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/329,624
; FILING DATE: 27-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/956,171
; FILING DATE: October 20, 1997

```

DB	1110	TTTATTCTCGATTTATAGAAAGACATTAATCTCTGTATACAAACAGCTTTATTTCCCTGAAAC	1169
Qy	778	TTTGAGACGGAGGTGTTTTTCAATACCAATCCCTTCGCCACGCTATGGAACGTCCTTC	837
Db	1170	TATGAATTAATTAAGTATAGACAATGGGAGTTTATCATCGGATGATCGGCTCT	1229
Qy	838	TTGATTCTTAATGCTACTCAAAATGGTACGTGTTAAGCTTGAGATTACTCAAAATCATATT	897
Db	1230	TTATTAGCGCGTGAAGGTGGTAAACAGTTTATTTAAATTTAAGTACAGGTGATGACGTTGTT	1289
Qy	898	TCAGCTCATGTTAACTCACTCAGCTGAGTTGGTAAAGTAAACGAGGATTTAGATATTTGTTAGT	957
Db	1290	GAATTTGTTCTCATACACAGAAATTTGGTACCTGTAAAGAAAGATTGATGCAACGAT	1349
Qy	958	CAGTCTGGTAGTGATTTAACTATCAGCTTTCAAATCCAACTTACCTTTATTTAGTCTTTAAAA	1017
Db	1350	GTTGAAGGTGGTAGCTGAAATTTTCATTCACCTCAATATATATGATGGATGCTTTAAAA	1409
Qy	1018	GCTATTAAAGTGAACAGPAAAAATTCATTTCTTATCACAGTTTCGACATTCACCCCTA	1077
Db	1410	GCAATCGATAATGATGAGGTTGAAGTTGAATTTCTCGGTACAAATGAAACCATTTATTCTA	1469
Qy	1078	ACACCGGCGATGAGGA	1094
Db	1470	AAACCAAAAGGTGACGA	1486

RESULT 17

US-09-815-242-8602

Sequence 8602, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: Fast-Seq for Windows Version 4.0

SEQ ID NO 8602

LENGTH: 1134

TYPE: DNA

ORGANISM: Staphylococcus aureus

FEATURE:

NAME/KEY: CDS

LOCATION: (1)...(1134)

US-09-815-242-8602

Query Match 20.5%; Score 233; DB 9; Length 1134;

Best Local Similarity 52.1%; Pred. No. 1.5e-38;



QY 538 ACTGACTCTCATCGTATGAGCAACGTTTAAATCACTTTGGACAATACTTTCAGCAGATTG 597  
 Db 538 ACTGATTCACACCGCTGCTGTAAAGAAAGTTGCGAGTTAGAAAGATGTTTCTGAAACAAA 597  
 QY 598 ATGGTAGTCTTCCAAAGTAAATCTTTGAGAGAAATTTTCAGCAGATTTTACAGATCATATT 657  
 Db 598 ATGTCATCATCTCCAGGTAGCGTTTACGTGAATTAATAAATTTATGTCTGCAATGAA 657  
 QY 658 GAGACGGTGTAGGATTTTCTTCCACCAAGCAAAATCTTTGAGAGTGAACACATTTCT 717  
 Db 658 GAAGACATGATATCTCTTCTGCTTCAACCAAGTTTATTTAAAGTTGGAATGTGAAC 717  
 QY 718 TTTTATACAGCGCTTTCAGAGAAATTTATCCGATACAGACCGTTTATTATATGACAGAA 777  
 Db 718 TTTATTTCTCGAATTTAGAGGACATTTATCTGATACACACGTTTATTTCCCTGAAAC 777  
 QY 778 TTTGAGCGGAGTGTGTTTCAATACCAATCCCTTCGCCACGCTATGGAACGTGCTTC 837  
 Db 778 TATGAATTAATTAAGTATAGACATGCGGAGTTTATCATGCGATGATGCTGCTCT 837  
 QY 838 TTGATTCTTAATGCTACTCAAAATGGTACTGTAAAGCTTGAGATTACTCAAAATCATATT 897  
 Db 838 TTATTAGCAGCGTGAAGTGGTAAATACGTTATTTAAATTAAGTACAGGTGATGACGTGTT 897  
 QY 898 TCAGCTCATGTTAACTCAGCTGAGTGGTAAAGTAAACGAGGATTTAGATATTGTTAGT 957  
 Db 898 GAATTTATCTTACATACCAAGAAATTTGTTACTGTAAAGAAAGAAAGTTGATGCAAAACGAT 957  
 QY 958 CAGTCTGGTAGTGAATTAATCACTCAGCTTCAATCAACTTACCTTATTGAGTCTTTAAA 1017  
 Db 958 GTTGAAGTGGTAGCTGAAATTTCAATCAACTTAAATATATATGATGATGCTTTAAA 1017  
 QY 1018 GCTATTAAGAGTGAACAGTAAAAATTCATTTCTTATCACCAGTTTCGACCAATTCACCTA 1077  
 Db 1018 GCAATCGATAATGATGAGTGAAGTTGAAGTTGAATCTTCCGTTACAAATGAACCAATTTATTCTA 1077  
 QY 1078 ACACCGAGCGATGAGGA 1094  
 Db 1078 AAACCAAAAGGTGACGA 1094

RESULT 19

US-10-282-122A-7900  
 ; Sequence 7900, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Travick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 ; FILE REFERENCE: ELTRA.034A  
 ; CURRENT APPLICATION NUMBER: US/10282,122A  
 ; CURRENT FILING DATE: 2003-02-20  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 7900  
 ; LENGTH: 1134  
 ; TYPE: DNA  
 ; ORGANISM: Staphylococcus aureus  
 US-10-282-122A-7900  
 Query Match 20.5%; Score 233; DB 16; Length 1134;  
 Best Local Similarity 52.1%; Pred. No. 1.5e-38;  
 Matches 571; Conservative 0; Mismatches 520; Indels 6; Gaps 2;  
 QY 1 ATGATTCAATTTTCAATTAATCGCACATTTATTTATTCATGCTTTAAATACAACTAAACGT 60  
 Db 1 ATGATGGAAATTCATTTTAAAGAGATTTATTTATACAAATTAATTAATGACACATTAATA 60  
 QY 61 GCTATTAGCAGTCAAAAATGCCATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 120  
 Db 61 GCTATTTCCACCAAGAACCAACATTTACCTATATTTAACTGGTATCAAAATCGATCGAAGAA 120  
 QY 121 ACAGAGTAACTTTTACAGGCTTAACGGTCAATATCAATTAATTAATTAATTAATTAATTAAT 180  
 Db 121 CATGAAGTTATATCAACTGGTTCAGACTCTGAAATTTCAATAGAAATCACTATTTCTTAAA 180  
 QY 181 AGT---AATGAAATGCTGGTTGCTTAATTTACCTCTCCAGGAGCTATTTTATTAGAGCT 237  
 Db 181 ACTGTAGATGCGAAGATATTGTCAATTTTTCAGAAACAGGCTAGTAGTACTTCTTGA 240  
 QY 238 AGTTTTTTTATTAATTTTCAAGTTTGGCAGATATTAGTATTAATTTTAAAGAAATTT 297  
 Db 241 CGATTTCTTTGATATTATAAAAAATTAACCTGGTAAAGATGTTTAAATTTATCTACAAAT 300  
 QY 298 GAACAAACCAAGTTGTTTAAACGAGTGTAAATCAGAGATTTACCTTTAAAGGAAAGAT 357  
 Db 301 GAACAATTTCCAGACATTTAAATTAATTAATTAATTTTAAAGTGGCTTAGAT 360  
 QY 358 GTTGACAGTATCTCTCGTCTCAAGAAAGTATCAACAGAAATTCCTTTTGAATTTTAAACAA 417  
 Db 361 CCAGATCAATATCTTTTATTAATTTCTAGAGATGACGCAATTCATTTGTCGGTA 420  
 QY 418 AAATTTTGAAGTCTATTATTGCTGAAAACAGCTTTTGCAGCCAGTTTACAAGAAAGTCT 477  
 Db 421 AAAGTGTAAAAACGTCGATTTGCAAAACGAAATTTTGCAGTGTCCACCTCCAGAAACACGC 480  
 QY 478 CCTATTTTAAACAGGATTCATATTGTATTAGTATCATTAAGATTTTAAAGCAGTAGCG 537  
 Db 481 CCAGTACTAACTGGTGTGAACCTGCTTATA---CAAGAAATTAATTAATGTCACAGCG 537  
 QY 538 ACTGACTCTCATCGTATGAGCAACGTTTAACTTTTGGCAATATCTTTCAGCAGATTG 597  
 Db 538 ACTGACTCAGACCGCTTGGCTGTGAAGAAAGTTGCAAGTTAGAGATGTTTCTGAAACAAA 597  
 QY 598 ATGGTAGTCTTCCAAAGTAAATCTTTGAGAGAAATTTTCAGCAGTATTTTACAGATGATTT 657  
 Db 598 AATGTCATCATTTCCAGGTAAAGCTTTAGCTGAATTTAAATTAATTAATTTATGCTGCAATGAA 657  
 QY \* 658 GAGACCGTTGAGGTATTTTCTTCCACCAAGCAAAATCTTGTTCAGAGTGAACACATTTCT 717  
 Db 658 GAAGACATTTGATATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 717  
 QY 718 TTTTATACACGCTCTTAGAAGGAAATTTATCCGATACAGACCGTTTATTTATTAATGACAGAA 777  
 Db 718 TTTATTTCTCGAATTTAAGAGGACATTTATCTGATACAAACGTTTATTTCTCCTGAAAC 777

Qy	778	TTTGAGAGCGGAGTGTGTTTTCAATACCCAAATCCCTTCGCCACGCTATGGAAACGTGCGCTTC	837
Db	778	TATGAAATTTAAATTTAAGTATATAGACAATGGGAGTGTATCATGCAATTGATCGTGCCTCT	837
Qy	838	TTGATTTCTAATGCTACTCAAAATGGTACTGTTAAAGCTTGAGATTACTCAAAATCATATT	897
Db	838	TTATTAGCAGTGAAGTGGTAAATACGTTATTAAATTAAGTCAGGTGATGACGTGGT	897
Qy	898	TCAGCTCATGTTAACTCACTGAGGTTGGTAAAGGTAAACGAGGATTTAGATATTGTTAGT	957
Db	898	GAATTATCTTCTACATCACCAGAAATGGTACTGTAAAGAAAGAAGTTGATGCAACCAT	957
Qy	958	CAGTCTCGGTAGTGATTTAACTATCAGCTTCAAATCCAACTTACCTTATTGAGTCTTTAAAA	1017
Db	958	GTTGAAGTGTGAGCCTGAAAATTCATTCAACTTAATATATGATGATGATCTTTAAAA	1017
Qy	1018	GCTATTTAAAGTGAACAGTAAAAATTCATTCTTATCAACAGTTGCAACATTCAACCTTA	1077
Db	1018	GCAATCGATTAATGATGAGGTTGAAGTTGAATCTTCGGTACAATGAACCATTTATCTA	1077
Qy	1078	ACACCGGGCGATGAGGA	1094
Db	1078	AAACCAAAAGGTGACGA	1094

## RESULT 20

```

US-09-815-242-4614
; Sequence 4614, Application US/09815242
; Patent No. US20020061569A1
GENERAL INFORMATION:
; APPLICANT: Haseelbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Esse
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/233,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 4614
LENGTH: 1128
TYPE: DNA
ORGANISM: Staphylococcus aureus
US-09-815-242-4614

```

```

Query Match      20.3%; Score 230; DB 9; Length 1128;
Best Local Similarity 51.9%; Pred. NO. 6.1e-38;
Matches 568; Conservative 0; Mismatches 6; Gaps 2;

OY      4  ATTCAAATTTTCAATTAATCGCATTATTATTATTCGTTTAAATACAACTAAACGTGCT 63
db      1  ATGGAATTCCTACTTAAAGAGATTATTATTATACCAATTAATGACACATTAAGAAGCT 60

```

Qy	64	ATTAGCACTAAAAATGCGCATTTCTTATCTTTTCATCAATAAAAAATTGAAGTCACCTTCTACA	123
Db	61	ATTTCCACCAAGAACAAACATATACCTATATTAACCTGGTATTTAAATTCGATGCAAAAAGAACAT	120
Qy	124	GGAGTAACTTTAAACAGGGTCTAAACGGTCAATATCAATTTGAAAACAATAATTCCTGTGAAGT	183
Db	121	GAAGTTATATTAACCTGGTTCAGACTCTCGAAATTTCAATAGAAATTCATATTTCTTAAAACT	180
Qy	184	---AATGAAAAATGCTGGTTTCTCTAAATTACCTCTCCAGGAGCTATTTTATTATAGAAGCTAGT	240
Db	181	GTAGATGCGGAAGATATTGTCAATATTTAGAAAACAGCTCAGTAGTACTTCTCTGGACGA	240
Qy	241	TTTTTTATTAATATATTTCAAGTTTCCAGATATTAGTATAAATGTTAAAGAAATTTGAA	300
Db	241	TTCTTTGTGTGATATATAAAAAAATTAACCTGGTAAAGATGTTAAATTTATCTACAAATGAA	300
Qy	301	CAACACCAAGTTGTTTTAAACAGGTGGTAAATCAGAGATTACCTTTAAAGAGAAAGATGTT	360
Db	301	CAATTCAGACATTAATTACATCAGGTCAATTTCTGAATTTAATTAAGTGGCTTAGATCCCA	360
Qy	361	GACCAGTATCCTCGTCTACAGAGAGTATCAACAGAAAATCCTTTTGAATTTTAAAAACAAAA	420
Db	361	GATCAATATCCTTTATTACCTCAAGTTCTAGAGATGACGCAATTCATTTGTCGGTAAACA	420
Qy	421	TTATTGAAGTCTATTATTGCTGAAACAGCTTTTTCAGCCAGTTTAAAGAAAAGTCGTCT	480
Db	421	GTGCTTAAAAACGTAATTGCAACAAACGAAATTTTTCAGTGTCCACCTCAGAAAACACGCCCA	480
Qy	481	ATTTTAAACAGGAGTTCCATATTTGTTAATGAAGTAATCATAAAGATTTTAAAGCAGTAGCCACT	540
Db	481	GTACTAATCGTGTGAATCGCTTATA--CAAGAAAATGAAATTAATATGCAACAGCACT	537
Qy	541	GACTCTCATCGTATAGCCAAAGTTTAACTCCTTTTGGACAATCTTCAGCAGATTTTGATG	600
Db	538	GACTCACACCGCTTGGCTGTGAAGAAAGTTTCGAGTTAGAAATGTTCTGAAAAACAAAAAT	597
Qy	601	GTAGTTCTTCCAAAGTAAATCTTTTGAGAGAAATTTTCAGCAGTATTTACAGATGATATTGAG	660
Db	598	GTCAATCATTCAGGTAAAGCTTTAGCTGAATTAATTAATTAATTAATGCTGCAATGAAGAA	657
Qy	661	ACGTTGAGGTAATTTTCTCCAAAGCAAAATCTTGTTCAGAAAGTGAACACATTTCTTTT	720
Db	658	GACATTTGATATCTTCTTGTCTTCAAAACCAAGTTTATTTAAAGTTCGAAATGTGAACCTTT	717
Qy	721	TATACACGCTCTTAGAAGGAATATCCGATACAGACGTTTATTAATATGACAGAAATTT	780
Db	718	ATTTCTCGATTATAGAGGACATATCTCTGATACAAACAGTTTATTTCCCTGAAAACTAT	777
Qy	781	GAGACGGAGGTTGTTTTCAATATACCAATCCCTTCGCCACGCTATGGAAGTGCGCTTCTTG	840
Db	778	GAATTTAAATTAAGTATAGCAATGGGAGTTTATCATGCGATTGATCGTGCACTTTTA	837
Qy	841	ATTTCTAATGCTACTCAAAATGGTACTGTTAAGCTTGAGATTACTCAAAATCATATTTCA	900
Db	838	TTAGCAAGTGAAGGTCGTAAATACGTTATTAATTAAGTACAGGTGATGACGTTGTGAA	897
Qy	901	GCTCATGTTAACTCACCTCAGGTTGGTAAAGTAAACGAGGATTTAGATATGTTAGTCAG	960
Db	898	TTATCTTCTACATCACAGAAATTTGTAATGTTAAAGAAAGATTTGATGCAACGATGTT	957
Qy	961	CTGGTAGTGAATTAATCTATCAGTTTCAATCCAATCTTATTGAGTCTTTTAAAGCT	1020
Db	958	GAAGGTGGTAGCCTGAAAAATTTCAATCACTCTAAATATGATGATGATGCTTTTAAAGCA	1017
Qy	1021	ATTTAAAGTGAAACAGTAAAAATTCATTTCTTATCACAGTTTCGACCAATTCACCTTAACA	1080
Db	1018	ATCGATAATGATGAGGTTGAGTTGAATTTCTTCGGTACAAAGAACCAATTTATCTTAAAA	1077
Qy	1081	CCAGCCGATGAGGA	1094
Db	1078	CCAAAAGGTGACGA	1091

```
RESULT 21
US-10-398-221-3763
; Sequence 3763, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3763
; LENGTH: 4736
; TYPE: DNA
; ORGANISM: Listeria monocytogenes 4b
; NAME/KEY: misc feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u
US-10-398-221-3763

Query Match 17.7%; Score 200.4; DB 16; Length 4736;
Best Local Similarity 51.5%; Pred. No. 1.7e-31;
Matches 565; Conservative 0; Mismatches 521; Indels 12; Gaps 4;

QY 2 TGATTCAATTTTCAATTAATCGCATTATTTATTCATGCTTTTAAATACAACTAAACGTTG 61
DB 1218 TCATGAAATTTGTTATTAGCGTGTGCTTGTCCAGCAGTCAATGAAGTTACTCGTG 1277
QY 62 CTATTAGCACTAAATGCAATTCCTATCTTCATCATATAAATTAAGTCACTTCTA 121
DB 1278 CCATCTCTGCAAGAACAGATTCCTCAATTTTAAACGGGGATAAATAGTCGTAATGATG 1337
QY 122 CAGGAGTAATTTTAAACAGGCTCTAAGGTCCTAATTAATCAATTAATCAATCTCTGTA 181
DB 1338 AAGGTGTAACATACTGCTAGTGTATCGATATTTCAATCGAGCATTTATTCATTA 1397
QY 182 GT---AATGAAATGCTGTTGCTTAATTAATCACTCTCCAGGAGCTATTTTATTAAGCTA 238
DB 1398 TTGAAATGATCAAGTAATTTAGAGTGAGAGTGTGTTGTTGGAATTTGACTTCAATCAA 1457
QY 239 GTTTTATTAATTAATTTTCAAGTTTCCAGATATTAGTATAAATGTTAAGAAATG 298
DB 1458 AATACCTTTGGCGATATTGTTGCTGTTTACCAGAAAGAAATGAGAAATTTGAAGTACTT 1517
QY 299 AACACACCAAGTTGTTTAAACAGTGGTAATTAATCAAGATTAACCTTTAAAGGAAAGATG 358
DB 1518 CTAACCTACCAACCAACATAGTTCTGGCAAGATCTCTTACATTAATAGCTTAGATC 1577
QY 359 TTGACCAAGTCTCTGCTTCAAGAGTATCAACGAAATCTTTGATTTTAAACAA 418
DB 1578 CAATGGAATATCTTAAATTTACCTGAAGTAACAGCGGAAACAAATTAATTTCAATTA 1637
QY 419 AATTATTGAAGTCTATTATTGCTGAAACAGCTTTTGCAGCAGTTTACAGAAAGTCTG 478
DB 1638 ATGTACTTTAAATAATTTGTTAGACAACTGTTTGTGCTGCTGCGATTGAAGTTCGTC 1697
QY 479 CTATTTTAAACAGGTTCTATATTGTTAAGTAATCAATAAGATTTTAAAGCAGTAGCGA 538
DB 1698 GAGTACTTACTGTTGTAATCTGATTAATCAAGAAATTA---CTAAGCGAGTTGCA 1754
QY 539 CTGACTCTCATGATGAGCCAGCTTTTAAATCACTTTGGACA---ATATTCAGCAGATT 595
DB 1755 CCGATAGTCACTGCTAGCTTTACGTTGAATACCTCTTGAAACAGACATTTGATGAAGAT 1814
QY 596 TGATGGTAGTCTTCTCAAGTAATCTTTGAGAGAAATTTTCAGCAGTATTTCAGATGATA 655

RESULT 22
US-10-398-221-9/c
; Sequence 9, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 684707
; TYPE: DNA
; ORGANISM: Listeria innocua
; NAME/KEY: misc feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-9

Query Match 17.5%; Score 198.8; DB 16; Length 684707;
Best Local Similarity 51.4%; Pred. No. 3e-30;
Matches 564; Conservative 0; Mismatches 522; Indels 12; Gaps 4;

QY 2 TGATTCATTTTCAATTAATCGCATTATTTATTCATGCTTTTAAATACAACTAAACGTTG 61
DB 267740 TCATGAAATTTGTTATTGAGCGTGTGCTTGTGTCAGAGCAGTCAATGAAGTTACTCGTG 267681
QY 62 CTATTAGCACTAAATGCAATTCCTATCTTTTCATCAATAAATTTGAAGTCACCTTCTA 121
```



Db 267680 CCATCTCTGCAAGACAAACGATTCCTCAATTTTAAACGGGATATAAAATAGTCGTAATGATG 267621  
 QY 122 CAGGAGTAACCTTTAAACAGGCTCTAACGGTCAATATCAATTTGAAACACACTATTCCTGTAA 181  
 Db 267620 AAGGTGTACTTTTAAACGGTAGGATCTGATATTTCCATCGAAGCAATTTATTCCTATTA 267561  
 QY 182 ---GTAATGAAATGCTGGTTGCTAATACCTCTCCAGGAGCTATTTTATTAGAAGCTA 238  
 Db 267560 TCGAAATGATGAAGTAATTTAGAGATTTGAAAGTTTGGTGGTATCGTTCTCAATCCA 267501  
 QY 239 GTTTTTTATTATATTTTCAAGTTTCCAGATATATAGTATAAATGTTAAAGAAATG 298  
 Db 267500 AGTACTTCGGTATATTTGTCGTGTTTACCTGAAGAAATGTCGAAATTTGAATTTACTA 267441  
 QY 299 AACCAACCAAGTTGTTTAAACAGTGTAAATCGAGATTACCTTAAAGCAAGAGATG 358  
 Db 267440 CAAATTTCAACTACATTTAGTTCTGGCCAGCTTCTTTTACACTAAACGGCTTAGATC 267381  
 QY 359 TTGACAGATATCTCGTCTCAAGAGATATCAACGAAATCCITTTGATTTTAAAAACAA 418  
 Db 267380 CAATGGAATACCTTAAATTACCAGAGTTTACTGATGGTAAAAATATTAATTTCTATTA 267321  
 QY 419 AATTATTGAAGTCTATTATTTGCTGAAACAGCTTTTGCAGCCAGTTTACAGAAAGTCGTC 478  
 Db 267320 ATGCTCTTAAAAACATTTAGACAAACCTGTTTTGCTGTTTCTGCTATCGAAGTTGCCC 267261  
 QY 479 CTATTTTAAACAGGAGTTCTATTTGATTTAAGTAATCATATAAGATTTTAAAGCAGTAGCA 538  
 Db 267260 CAGTCTTACAGGTGTAAGTATTTTAAAGATTTTAAAGCAGTAGCA 267204  
 QY 539 CTGACTCTCATGATGAGCCAACTTTAATCACTTTGGA---CAATACCTTACAGAGAT 595  
 Db 267203 CCGATAGTCACTGCTAGCATTTACGTGAAATTTCCACTTTGAAACAAATATTGACGAAGAT 267144  
 QY 596 TGATGTAGTCTTCCAGTAATCTTTGAGAGATTTTTCAGCAGTATTTACAGATGATA 655  
 Db 267143 ACACATTTGTTTTCAGGAAAGAGCTTAGCTGAACTAAATAAATTTTACGAGATGCA 267084  
 QY 656 TTGAGACCGTTGAGGTATTTTCTCACCAGCCAAATCTTGTTCAGAGTGAACACATTT 715  
 Db 267083 GTGAATCCATTGAATGACACTCGCAATATCAATCTTATTTAAATTAAGATTTAT 267024  
 QY 716 CTTTTTATACAGCCTCTTAGAGAAATTTTCCGATACAGACGTTTATTAATGACAG 775  
 Db 267023 TATTTTATTTCTGTTTACTGGAAGGTAGTTTACCCAGATACATCACGCTTAATTTCAACTG 266964  
 QY 776 AATTGAGACGGAGGTGTTTCAATACCAATCCCTTCCGACGCTATGAAAGTGCCT 835  
 Db 266963 ATACTAATCAGAACTTGTCAATTTCAAGAGCATTTTAAAGCGATTTGACCGTGTCT 266904  
 QY 836 TCTTGATTTCTAATGCTACTCAAAATGGTACTGTTTAAAGCTTTGAGATTTACTCAAAAT---C 892  
 Db 266903 CCTACTTGTCTCGCAAAACCGTAATACGTTTATTAATTAATGACCTTCAATGGAC 266844  
 QY 893 ATATTTACAGTCACTTAACTCACTGAGTTGTTGTAAGTAAACAGAGATTTAGATTTG 952  
 Db 266843 AAGTTGAAGTATCCTCTAACTCTCTGGAAGTTGGGAATGTTTTCGGAATGCTCTTAGTC 266784  
 QY 953 TTAGTCACTGTTAGTATTAATCACTATCAATCCAACTTCAATCTTACCTTATTCAGTCTT 1012  
 Db 266783 AAAGCTTTACTGTTGAGAGAAATTAATATCTTTTACGGTAAATATCATGATGATGAT 266724  
 QY 1013 TAAAGCTATTAAAGTGAACAGTAAATAATTTCTTTATATCAACGATTCGACCATTTCA 1072  
 Db 266723 TCGCGCATTTTGAAGCGGATGATATTTCAATCTCTCTCTGCGCACAATGAGACCTTTTCG 266664  
 QY 1073 CCCTAACACACAGCGATG 1090  
 Db 266663 TACTTCGACCAAAAGATG 266646

RESULT 23

US-10-398-221-2058

; Sequence 2058, Application US/10398221  
 ; Publication No. US20040018514A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KUNST, Frederik  
 ; APPLICANT: GLASER, Philippe  
 ; TITLE OF INVENTION: Listeria innocua, genome and applications  
 ; FILE REFERENCE: 344 702 - US  
 ; CURRENT APPLICATION NUMBER: US/10/398,221  
 ; PRIOR FILING DATE: 2003-03-27  
 ; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061  
 ; PRIOR FILING DATE: 2001-10-04  
 ; PRIOR APPLICATION NUMBER: FR 00/12 697  
 ; NUMBER OF SEQ ID NOS: 4025  
 ; SOFTWARE: Patent in version 3.0  
 ; SEQ ID NO 2058  
 ; LENGTH: 3011208  
 ; TYPE: DNA  
 ; ORGANISM: Listeria innocua  
 ; US-10-398-221-2058

Query Match 17.5%; Score 198.8; DB 16; Length 3011208;  
 Best Local Similarity 51.4%; Pred. No. 5.6e-30;  
 Matches 564; Conservative 0; Mismatches 522; Indels 12; Gaps 4;

QY 2 TGATTCAATTTTCAATTAATCGCACATTTATTTATTCATGCTTTAAATCAACTAAACGTG 61  
 Db 1866 TCATGAATTTGTTTATTGAGCGTGATCGTCTTGCCAAAGCAGTCAATGAAGTTACTCGTG 1925  
 QY 62 CTATTAGCACTTAAATAATGCCATTTCTTATTTCTCATCAATAAAAAATTTGAAGTCACTTCTA 121  
 Db 1926 CCACTCTGCAAGAACACGATTTCCATTTTAAACGGGATATAAATAGTCGTAATGATG 1985  
 QY 122 CAGGAGTAACCTTTAAACAGGCTTAAACGGTCAATATCAATTTGAAACACACTATTCCTGTAA 181  
 Db 1986 AAGGTGTACTTTAAACGGTAGCGATTTCTGATATTTCCATCGAAGCATTTTATTCATTA 2045  
 QY 182 ---GTAATGAAATGCTGGTTTGTCTAATTAACCTCTCCAGGAGCTATTTTATTAGAAGCTA 238  
 Db 2046 TCGAAATGATGAAGTAATTTGTAAGTTGAAAGTTTGGTGGTATCGTCTTCAATCCA 2105  
 QY 239 GTTTTTTATTATTAATTTTCAAGTTTGGCAGATATTTAGTATAAATGTTTAAAGAAATG 298  
 Db 2106 AGTACTTCGGTATTTGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2165  
 QY 299 AACACACCAAGTTGTTTTAAACAGTGTAAATCAGAGATTTACCTTAAAGAGAAAGATG 358  
 Db 2166 CAAATTAACAACTTAACATTTAGTTTCTGCGCAAGCTTCTTTACACTAAACGGCTTAGATC 2225  
 QY 359 TTGACCACTATCTCGTCTCAAGAGATATCAACAGAAATCTTTTCAATTTTAAACAA 418  
 Db 2226 CAATGGAATACCTTAAATTAACAGAGATTTACTGATGTAATAATTTAAATTTCTATTA 2285  
 QY 419 AATTATTGAAGTCTATTTTGTGAAACAGCTTTTGCAGCCAGTTTACAGAAAGTCGTC 478  
 Db 2286 ATGCTCTTAAAAACATTTATAGACAAACCTGTTTTGCTGTTTCTGCTATCGAAGTTGCCC 2345  
 QY 479 CTATTTTAAACAGGAGTTCTATTTGTAATTAAGTATCATAAAGATTTTAAAGCAGTAGCGA 538  
 Db 2346 CAGTCTTACAGGTGTTTAACTGGATTTAT---TAAAGATAACAACTTAGCGCAGTTGCAA 2402  
 QY 539 CTGACTCTCATCTGATGAGCCAACTGTTTAACTCACTTTGGA---CAATACCTTACAGCAGAT 595  
 Db 2403 CCGATAGTCACTCTAGCATTACGTAATTTCCACTTGAACAAATATTGACGAGAT 2462  
 QY 596 TGATGTAGTCTTCCAGTAATCTTTGAGAGAAATTTTTCAGCAGATTTTACAGATGATA 655  
 Db 2463 ACAACATTTGTTATTCAGGAAAGCTTAGCTGAATTAATAAATTTTACAGCAGTGA 2522  
 QY 656 TTGAGACCGTTGAGGTATTTTCTCACCAGCCAAATCTTGTTCAGAGTGAACACATTT 715  
 Db 2523 GTGAATCCATTGAATGACACTCGCAATTAATCAATCTTATTTAAATTAAGATTTAT 2582





```
US-09-974-300-1554
; Sequence 1554, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1554
; LENGTH: 1072
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1554

Query Match          9.3%; Score 105.8; DB 9; Length 1072;
Best Local Similarity 47.3%; Pred. No. 4.8e-12;
Matches 456; Conservative 0; Mismatches 497; Indels 12; Gaps 4;

QY 61 GCTATTAGCACTAAATAATGCCATTCCTATTCTTTCATCAATAAATAATGAAAGTCACATTCT 120
DB 58 GCGGTTCTTCAAGAACGACGATTCGATCTTAACCGGTATTAAATCGTGGCCTCGAT 117
QY 121 ACAGAGTAACCTTTAAACAGGGTCTAACCGGTCAATATCAATATGAAAAACACTATTCTCTGTA 180
DB 118 GAAGGGGCTCTCTCACAGGCGAGGATTCGATATTCGATTTGATTTGATTCGCGAAA 177
QY 181 ACTAATGAAT--GCTGGTTGCTAATTAACCTCCAGGAGCTATTTTATTAGAGCT 237
DB 178 GAAGACGGCGATTAGAGATCGTGAATTTGAACGCCGCGAGCATTTGCTTCAAGCC 237
QY 238 AGTTTTTTTAAATATATTATTTCAAGTTTGCAGATATTAGTATTAATGTTTAAAGAAAT 297
DB 238 GCTTTTTCAGTGAATTTGTCAAAGCTGCGATGTCAACAGTGAATTCAGGTTCAA 297
QY 298 GAACAACCAAGTTGTTTAAACAGTGGTAAATCAGAGATTAACCTTTAAAGGAAAGAT 357
DB 298 AATCAATACTTAAACGATCATCGCTCCGCAAGCAGAGTTTAACTTAAACGGTTTGGAT 357
QY 358 GTTGACCAAGTATCCGCTCAAGAGTATCAACAGAAATCCCTTGAATTTTAAACA 417
DB 358 GAAGCGAATATCCGCTTTGCGCAAAATGAAGAGCATCACGCTTTTCAAAATCCGACC 417
QY 418 AAATTTATGAAGTCTATTATTGCTGAAACAGCTTTTGACGCGAGTTTACAAGAAAGTCGT 477
DB 418 GATCTGCTGAAACAACTGATCCGCAACCGTTTGTGAGTGTCCACCTCAGAAACAGC 477
QY 478 CTTATTAAACAGGAGTTCATATTGTAATTAAGTAATCAATAAGATTTTAAAGCAGTAGCG 537
DB 478 CCAATCTTCACAGGTGTAATGTAATGTCACTGGCGGT---GAATTAATATGCACTGCA 534
QY 538 ACTGACTCTCATCGTATGAGCCAAAGTTTATCACTTTTGGACAACTTTCAGCAGAT--- 594
DB 535 ACGGATATGTCATCGCTTTCGCGTAAGGAAAGCTAAGCTCGACATTAACGAAGACAGTTCA 594
QY 595 TTGATGGTAGTCTTCCAAAGTAAATCTTTGAGAGAAATTTTACGAGTATTTACAGATGAT 654
DB 595 TACATGTGCTCATCCAGGAAAAGCTTAACCGAGCTCAGCAAAATCTTGTATGACCAT 654
QY 655 ATTGAGACGGTTGAGGTATTTTCTCAACAAGCCAAATCTTGTTCAGAAAGTGAACACAT 714
DB 655 CAGGAGCTTGTGATTTGATTTGATTTACCGAAACACAAAGTGTCTGTTTAAACAAACCGTT 714
QY 715 TCTTTTATACGGCTCTTAGAAGGAAATTTATCCGATACAGACCGCTTTTATTATGACA 774

US-10-282-122A-17164
; Sequence 17164, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 17164
; LENGTH: 1206
; TYPE: DNA
; ORGANISM: Clostridium difficile
US-10-282-122A-17164
```



QY 577 GACAATCTTACGAGATTGAGGTAGTCTTCCAGTAATAATCTTGGAGAGAAATTTTCA 636  
DB 571 GTTGATAATGATAATACGATAAATGCTGTATTTCCAGGGAACAATTAATGAAATTTTCA 630  
QY 637 CGAGTATTTACAGATGATATTGAGACCGTTGAGGTATTTTCTCACCAAGCCAAATCTTG 696  
DB 631 AAAATTTTAGAAGAAGAAATGAAATGTTAATATACATTTTACTCTTAAATCATATTTTA 690  
QY 697 TTCAGAAGTGAAACACATTTCTTTTATACACCGCTCTTAGAGGAAATTTATCCGATACA 756  
DB 691 TTTAGTATAGGAGAAACAAAATAATTTCAAGATTATTAGAGGAGAAATTTATTAGCTAT 750  
QY 757 GACCGTTTATTAATGACAGAAATTTGAGACGGAGGTGTTTTCATACCAATCCCTTCGC 816  
DB 751 AAGCTATAATACCTGAAGAATTTAACTTTAAATAAATAGCTTAAAGATCAGAGCTTTTA 810  
QY 817 CACGCTATGGAACGCTCTTCTGATTTCTTAATGCTACTCAAAATGGTACTGTTAAGCTT 876  
DB 811 AATCTATAGAAAGACCTCTCTTATGGCTAAGAGGAATACTAAATTTAGTTAAATTT 870  
QY 877 GAGATTACTCAAAATCATATTTTCACTCATGTAACTCACCTGAGGTGGTAAAGTAAAC 936  
DB 871 GATTTCTCAGATGATAAATTTGTAATAACATCAAAATCTC---AATTAGGAATGTTTGA 927  
QY 937 GAGGATTTAGATATTTGTTAGTCAGTCTGTGTAGTATGATTTAACTATCAGCTTCAATCCAAT 996  
DB 928 GAAGAATTAAGAATTTGTTATGCA---GGAGAAGATTTTCAAAATTTGCAATTTAAATTTCAAAA 984  
QY 997 TACCTTATGAGTCTTTAAAGCTATTAAAGCTGAAACAGTAAATAATTCATTTCTTATCA 1056  
DB 985 TATCTTTTAGATGTTGTAAGAACTATGGAAGATAATAGAGTTGTTATTAGAAATTTTCAAGT 1044  
QY 1057 CCAGTTTCGACCAATTCACCCCTACA 1080  
DB 1045 AGTGAAGTCTTGTATTATATAAA 1068

## RESULT 29

US-10-671-403-173  
; Sequence 173, Application US/10671403  
; Publication No. US20040038289A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Donnell, Michael E.  
; APPLICANT: Yuzhakov, Alexander  
; APPLICANT: Yurieva, Olga  
; APPLICANT: Jeruzalmi, David  
; APPLICANT: Bruck, Irina  
; APPLICANT: Kuriyan, John  
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT  
; FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND  
; TITLE OF INVENTION: USE THEREOF  
; FILE REFERENCE: 22221/1030  
; CURRENT APPLICATION NUMBER: US/10/671,403  
; PRIOR APPLICATION NUMBER: 2003-09-25  
; PRIOR FILING DATE: 2000-11-21  
; PRIOR APPLICATION NUMBER: 60/143,202  
; PRIOR FILING DATE: 1997-04-08  
; PRIOR APPLICATION NUMBER: 08/823,407  
; PRIOR FILING DATE: 1997-04-08  
; PRIOR APPLICATION NUMBER: 09/057,416  
; PRIOR FILING DATE: 1998-04-08  
; NUMBER OF SEQ ID NOS: 212  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 173  
; LENGTH: 992  
; TYPE: DNA  
; ORGANISM: Bacillus stearothermophilus  
US-10-671-403-173

Query Match 7.8%; Score 88; DB 16; Length 992;  
Best Local Similarity 46.5%; Pred. No. 2.4e-08;

Matches 390; Conservative 0; Mismatches 440; Indels 8; Gaps 3;  
QY 214 CCAGAGCTATTTTATTAGAACTAGTGTGTTTATTAATATTTTCAAGTTTGCAGAT 273  
DB 73 CCGGGAGCATCGTACTGACGCGCTTTTCTCTGAAATCGTGAAAAAATCGCGCAA 132  
QY 274 ATTAGTATAATGTTTAAAGAAATTTGAACAAACCAAGTTGTTTAAACAGTGGTAAATCA 333  
DB 133 CAAACGGTGGAAATCGAAACCGAAGACAACTTTTGTACGATCATCCGCTCGGGGCACTCA 192  
QY 334 GAGATTACTTTAAAGAGAAAGATGTTGACCAAGTATCTCTGCTACAAAGAACTATCAACA 393  
DB 193 GAATTCGCCCTCAATGGCTAAACGCCGACGAAATATCCGCCCTCGCCCAATTTGAAGAA 252  
QY 394 GAAATCTTTGATTTTAAACAAATTTAAGATCTATATTGCTGTGAACAGCTTTT 453  
DB 253 GAAACGGTGTTCAAATCCCGCTGATTTATTGAAACCGTGATTCGGCAACCGGTGTTTC 312  
QY 454 GCAGCCAGTTTACAAGAAAGTCTGCTATTTTACAGGAGTTCATATTGTTATTAGTAAT 513  
DB 313 GCGGTTTCTACATCGGAAACCGGCCAATCTTGACAGGTGCAACTGGAAAGT---TGAA 369  
QY 514 CATAAAGATTTTAAAGCAGTAGCAGTACTCTCATCTGATGAGCAACGTTTAAATCACT 573  
DB 370 CATGCGAGCTGTGTCACAGCGACGACAGTCACTCGTTAGCCATGGCAAGTGA 429  
QY 574 TTGGACA---ATACCTTACAGCAGATTTGATGGTGTCTTCCAAAGTAAATCTTTTGAGAGAA 630  
DB 430 ATTGAGTCGGAATAATGAAGTATCATACAACGTCGTCATCCCTGGAAAAAGTCTTAATGAG 489  
QY 631 TTTTCAGCAGTATTTACAGATGATATTGAGACCGTGTAGGTATTTTCTCACCAGCCAA 690  
DB 490 CTCAGAAAAATTTTGGATGACGGCAACCAACCGGTGGACATCGTCATGACAGCCAAATCAA 549  
QY 691 ATCTGTTTCAAGAGTGAACACATTTCTTTTATACACGCTCTTATAGAGGAAATTTATCCC 750  
DB 550 GTGCTATTTAAGCGGAGACCTTCTCTTTTCCCGGCTGCTTGAACGCAACTATCG 609  
QY 751 GATACAGACCGTTTATTAATGACAGAAATTTGAGACGGAGGTGTTTCAATACCAATCC 810  
DB 610 GAGACGGCCGCTTGATTTCAACAGAAAGCAACGACCATGATCGTCAATGCAAAAGAG 669  
QY 811 CTTGCCACGCTATGGAACGTCCTTCTTGATT---TCTAATGCTACTCAAAATGTACTG 868  
DB 670 TTTCTGAGGCAATGACGCGGCTCTTGTGCTGAGAGGAAAGAACACGTTG 729  
QY 869 TTAGCTTTGAGTACTCAAAATCATATTTTACGCTCATGTTAACTCACTGAGGTGGTA 928  
DB 730 AAACGTGACGCGCTTCTGGAGGAATGCTCGAAATTTCTCGATTTCTCCGAGATCGGA 789  
QY 929 AGGTAACGAGGATTTAGATATTGTTAGTCACTGTTGTTAGTATTAATCACTATCAGCTTCA 988  
DB 790 AAGTGACGGAGCAGCTGCAAAACGGAGTCTCTTGAAGGGGGAAGAGTTGAACATTTCTGTTCA 849  
QY 989 ATCCAATCTACTTATGAGTCTTTTAAAGCTATTTAAAGTGAACAGTAAAAATTTCA 1046  
DB 850 GCGCAAAATATATGATGACGCGTTGCGGGCGCTTGTATGGAACAGACATTTTCAATCA 907

## RESULT 30

US-10-671-419-173  
; Sequence 173, Application US/10671419  
; Publication No. US20040038290A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Donnell, Michael E.  
; APPLICANT: Yuzhakov, Alexander  
; APPLICANT: Yurieva, Olga  
; APPLICANT: Jeruzalmi, David  
; APPLICANT: Bruck, Irina  
; APPLICANT: Kuriyan, John  
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT  
; FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND  
; TITLE OF INVENTION: USE THEREOF



```
; FILE REFERENCE: 22221/1030
; CURRENT APPLICATION NUMBER: US/10/671,419
; CURRENT FILING DATE: 2003-09-25
; PRIOR APPLICATION NUMBER: US/09/716,964A
; PRIOR FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 60/143,202
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 08/823,407
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 09/057,416
; PRIOR FILING DATE: 1998-04-08
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 173
; LENGTH: 992
; TYPE: DNA
; ORGANISM: Bacillus stearothermophilus
US-10-671-419-173

Query Match      7.8%; Score 88; DB 16; Length 992;
Best Local Similarity 46.5%; Pred. No. 2.4e-08;
Matches 390; Conservative 0; Mismatches 440; Indels 8; Gaps 3;

QY      214  CCAGGAGCTATTTATTAGAGCTAGTCTTTTATTAAATTAATTTTAAAGTGGCAGAT 273
Db      73  CCGGGAGCATCGTACTGACGGCGCGCTTTTCTGAAATCGTGAATAATCGTGAATAATCGCGCAA 132

QY      274  ATTAGTATAAATGTTTAAAGAAATTTGAACCAACCAAGTTGTTTAAACAGTGTAAATCA 333
Db      133  CAAACGGTGAATTCGAAACCGGAAGACAATCTTTTGAAGATCATCCGCTCGGGGCACTCA 192

QY      334  GAGATTACCTTAAAGAAAGATGTTGACCATGCTCTCGTCTACAGAAAGTATCAACA 393
Db      193  GAATTCGGCTCAATCGGCTTAAACGGCGACGAATATCCGCGCTCGCGCAATTTGAAGAA 252

QY      394  GAAATTCCTTTGATTTTAAACCAAAATTTATCAAGTCTATTATTGCTGAAACAGCTTTT 453
Db      253  GAAAACGTGTTTCAATCCCGGCTGATTTATTGAAAACCGTGATTCGGCAACCGGTGTC 312

QY      454  GCAGCCAGTATTACAGATGATATTGAGACCGTGTGAGGTATTCTTCCACCAAGCCAA 690
Db      430  ATTGAGTCGGAATATGAAGTATCATACACGTCGTCATCCCTGGAAAAGTCTTAAATGAG 489

QY      631  TTTTCAGCAGTATTACAGATGATATTGAGACCGTGTGAGGTATTCTTCCACCAAGCCAA 690
Db      490  CTCAGCAAAATTTTGGATGACGGCAACCAACCGGTTGGACATCGTCATGACAGCCAAATCA 549

QY      691  ATCTTTGTCAGAGTGAACATTTCTTTTATACACCGCTCTTAGAAGGAAATATCC 750
Db      550  GTGCTATTTAAGCGCGAGCATCTCTCTTTTCCCGGCTGCTTGACGGCAACTATCCG 609

QY      751  GATACAGACCGTATTAAATGACAGAAATTTGAGACGGAGGTGTTTTCATACCAATCC 810
Db      610  GAGACGGCCCGTGTGATTTCCACAGAAAGCAACGACCATGTCGTCAATTCGAAAAGAG 669

QY      811  CTTGCCAGCTATGGAAGCGCTCTTTGATTTTCAAGTCTCATGTTAACTCAGCTGAGTGTG 868
Db      670  TTTCTGACGAGCAATCGACCGCGCTCTTGTGCTCGAGAAGGAAACAAAGTGTG 729

QY      869  TTAAGCTGAGTACTCAAAATCATATTTTCAGCTCATGTTAACTCAGCTGAGTGTGTA 928
Db      730  AAACTGACGAGCTTCTGGAGGAATGCTCGAAATTTCTCGATTTCTCGAGATCGGGA 789

QY      929  AGGTAAACGAGATTTAGATATTTGATGCTAGTCTGAGTGTGATTTAACTATCAGCTTCA 988
Db      931  TTTTCAGCAGTATTACAGATGATTTGAGACCGGTGAGGTATTTTCTCACCACCAAGCCAA 690

; FILE REFERENCE: 22221/1030
; CURRENT APPLICATION NUMBER: US/10/671,419
; CURRENT FILING DATE: 2003-09-25
; PRIOR APPLICATION NUMBER: US/09/716,964A
; PRIOR FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 60/143,202
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 08/823,407
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 09/057,416
; PRIOR FILING DATE: 1998-04-08
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 173
; LENGTH: 992
; TYPE: DNA
; ORGANISM: Bacillus stearothermophilus
US-10-670-844-173

Query Match      7.8%; Score 88; DB 16; Length 992;
Best Local Similarity 46.5%; Pred. No. 2.4e-08;
Matches 390; Conservative 0; Mismatches 440; Indels 8; Gaps 3;

QY      214  CCAGGAGCTATTTATTAGAGCTAGTCTTTTATTAAATTAATTTTAAAGTGGCAGAT 273
Db      73  CCGGGAGCATCGTACTGACGGCGCGCTTTTCTGAAATCGTGAATAATCGTGAATAATCGCGCAA 132

QY      274  ATTAGTATAAATGTTTAAAGAAATTTGAACCAACCAAGTTGTTTAAACAGTGTAAATCA 333
Db      133  CAAACGGTGAATTCGAAACCGGAAGACAATCTTTTGAAGATCATCCGCTCGGGGCACTCA 192

QY      334  GAGATTACCTTAAAGAAAGATGTTGACCATGCTCTCGTCTACAGAAAGTATCAACA 393
Db      193  GAATTCGGCTCAATCGGCTTAAACGGCGACGAATATCCGCGCTCGCGCAATTTGAAGAA 252

QY      394  GAAATTCCTTTGATTTTAAACCAAAATTTATCAAGTCTATTATTGCTGAAACAGCTTTT 453
Db      253  GAAAACGTGTTTCAATCCCGGCTGATTTATTGAAAACCGTGATTCGGCAACCGGTGTC 312

QY      454  GCAGCCAGTATTACAGATGATATTGAGACCGTGTGAGGTATTCTTCCACCAAGCCAA 690
Db      430  ATTGAGTCGGAATATGAAGTATCATACACGTCGTCATCCCTGGAAAAGTCTTAAATGAG 489

QY      631  TTTTCAGCAGTATTACAGATGATATTGAGACCGTGTGAGGTATTCTTCCACCAAGCCAA 690
Db      490  CTCAGCAAAATTTTGGATGACGGCAACCAACCGGTTGGACATCGTCATGACAGCCAAATCA 549

QY      691  ATCTTTGTCAGAGTGAACATTTCTTTTATACACCGCTCTTAGAAGGAAATATCC 750
Db      550  GTGCTATTTAAGCGCGAGCATCTCTCTTTTCCCGGCTGCTTGACGGCAACTATCCG 609

QY      751  GATACAGACCGTATTAAATGACAGAAATTTGAGACGGAGGTGTTTTCATACCAATCC 810
Db      610  GAGACGGCCCGTGTGATTTCCACAGAAAGCAACGACCATGTCGTCAATTCGAAAAGAG 669

QY      811  CTTGCCAGCTATGGAAGCGCTCTTTGATTTTCAAGTCTCATGTTAACTCAGCTGAGTGTG 868
Db      670  TTTCTGACGAGCAATCGACCGCGCTCTTGTGCTCGAGAAGGAAACAAAGTGTG 729

QY      869  TTAAGCTGAGTACTCAAAATCATATTTTCAGCTCATGTTAACTCAGCTGAGTGTGTA 928
Db      730  AAACTGACGAGCTTCTGGAGGAATGCTCGAAATTTCTCGATTTCTCGAGATCGGGA 789

QY      929  AGGTAAACGAGATTTAGATATTTGATGCTAGTCTGAGTGTGATTTAACTATCAGCTTCA 988
Db      931  TTTTCAGCAGTATTACAGATGATTTGAGACCGGTGAGGTATTTTCTCACCACCAAGCCAA 690
```

Db 490 CTCAGCAAAATTTTGGATGACGGCAACCCCGGTGGACATCGTCATGACAGCCCAATCAA 549  
 Qy 691 ATCTGTTTCAAGTGAACACATTTCTTTTATACAGCCTCTTAGAAGGAAATTTATCCC 750  
 Db 550 GTGCTATTAAAGCCGAGCACCTTCTTCTTTTCCCGCTGCTTGACGGCAACTATCOG 609  
 Qy 751 GATACAGACCGTTTATTAATGACAGATTTGAGACGGAGGTTGTTTTCAATACCAATCC 810  
 Db 610 GAGACGGCCGCTTGAATTCACAGAAAGCAACGACCATGATCGTCAATGCAAAAGAG 669  
 Qy 811 CTTCGCCAGCTATGAAAGTCTTCTTGTATT--TCTAATGCTACTCAAAATGCTACTG 868  
 Db 670 TTTCTGACGCAATGACCGAGCTCTTCTGCTTGTGCTGGAAGGAGCAACGCTTGTG 729  
 Qy 869 TTAAGCTTTGAGATTACTCAAAATCATATTTTCAGCTCATGTTAACTCACTGAGGTTGTA 928  
 Db 730 AAATGACGACGCTTCTCGGAGGATGCTCGAAATTTCTCGATTCTCCGAGATCGGA 789  
 Qy 929 AGGTAAACGAGATTTAGATATTTAGTACGCTCTGTTAGTGTATTAATCAATCAGCTTCA 988  
 Db 790 AAGTGACGAGCAGCTGCAACGAGTCTCTTTGAAAGGAGGAGTTGAACATTTTCGTTCA 849  
 Qy 989 ATCCAACTTACCTTATGAGTCTTTAAAGCTATTAAAGCTATTAAGTGAACAGTAAATTC 1046  
 Db 850 GCGGAAATATATGATGACGCTTGGCGGCTTGTGGAACAGACATTTCAAATCA 907

RESULT 32  
 US-10-671-134-173  
 ; Sequence 173, Application US/10671134  
 ; Publication No. US20040043415A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: O'Donnell, Michael E.  
 ; APPLICANT: Yuzhakov, Alexander  
 ; APPLICANT: Yurieva, Olga  
 ; APPLICANT: Jeruzalmi, David  
 ; APPLICANT: Bruck, Irina  
 ; APPLICANT: Kuriyan, John  
 ; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT  
 ; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND  
 ; TITLE OF INVENTION: USE THEREOF  
 ; FILE REFERENCE: 2221/1030  
 ; CURRENT APPLICATION NUMBER: US/10/671,134  
 ; PRIOR FILING DATE: 2003-09-25  
 ; PRIOR APPLICATION NUMBER: US/09/716,964A  
 ; PRIOR FILING DATE: 2000-11-21  
 ; PRIOR APPLICATION NUMBER: 60/143,202  
 ; PRIOR FILING DATE: 1997-04-08  
 ; PRIOR APPLICATION NUMBER: 08/823,407  
 ; PRIOR FILING DATE: 1997-04-08  
 ; PRIOR APPLICATION NUMBER: 09/057,416  
 ; PRIOR FILING DATE: 1998-04-08  
 ; NUMBER OF SEQ ID NOS: 212  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 173  
 ; LENGTH: 992  
 ; TYPE: DNA  
 ; ORGANISM: Bacillus stearothermophilus  
 US-10-671-134-173

Query Match 7.8%; Score 88; DB 16; Length 992;  
 Best Local Similarity 46.5%; Pred. No. 2.4e-08;  
 Matches 390; Conservative 0; Mismatches 440; Indels 8; Gaps 3;  
 Qy 214 CCAGGAGCTATTATTAGAACCTAGTTTATTAATTAATTTTCAAGTTTGCAGAT 273  
 Db 73 CCGGGAGCATCGTACTCGAGCGGCTTTTCTCTGAAATCGTGAATACTGCGGCA 132  
 Qy 274 ATTAGTATAAATGTTAAAGAAATTTGAACACCAAGTTGTTTAAACAGTGGTAAATCA 333  
 Db 133 CAAACGGTGAATCGAAACGGAAGACACTTTTTCAGCATATCCGCTCGGGGCACTCA 192

Qy 334 GAGATTACCTTAAAGGAAAGATGTTGACAGATATCTCTGCTTACAGAAGATATCAACA 393  
 Db 193 GAATTCGCGCTCAATGGGCTAAACGCGCGAGAAATATCCGCGCTGCGCNCNAATTTGAAGAA 252  
 Qy 394 GAAATCCCTTTGATTTTAAACAAATAATTTTGAAGTCTATTATTGCTGAAACAGCTTTT 453  
 Db 253 GAAACGCTGTTTCAATCCCGCTGATTTATTGAAACCGGTGATTCGGCAACCGGTGTTT 312  
 Qy 454 GCAGCAGTTTACAGAAAGTCTCTATTTTTAAACAGAGTTTCATATTGTTTAAAGTAAAT 513  
 Db 313 GCGGTTTCTACATCGGAAACGCGCCAATCTTGACAGGTGTCAACTGGAAGT---TGAA 369  
 Qy 514 CATAAAGATTTTAAAGCAGTAGGACACTCTCATCTGATGAGCAACAGTTTATCACT 573  
 Db 370 CATGCGAGCTTGTCTGCACAGCGACCGACAGTATCGCTTAGCCATCGGCAAGTGA 429  
 Qy 574 TTGACA---ATACTTTCAGCAGATTTGATGTTGATGTTCTTCCAAAGTAAATCTTTTGAGAGAA 630  
 Db 430 ATTGAGTCGGAATATGAATATCATACAACTGCTCATCCCTGGGAAAGTCTTAAATGAG 489  
 Qy 631 TTTTCAGCAGTATTACAGATGATTTAGAGACCGTTGAGGTTTCTTCAACCAAGCCAA 690  
 Db 490 CTCAGCAAAATTTTGGATGACGGCAACCAACCGGTGGACATCGTCATGACAGCAATCAA 549  
 Qy 691 ATCTGTTTCAGAGTGAACACATTTCTTTTATACAGCCTCTTGAAGGAAATTTATCCC 750  
 Db 550 GTGCTATTAAAGCGGAGCACCTTCTCTTTTCCCGCTGCTTGAAGGCAACTATCCG 609  
 Qy 751 GATACAGACCGTTTATTAAATGACAGAAATTTGAGACGGAGGTTGTTTCAATACCAATCC 810  
 Db 610 GAGACGGCCGCTTGAATTCACAGAAAGCAACGACCATGATCGTCATGCAAAAGAG 669  
 Qy 811 CTTCGCCAGCTATGGAACGCTTCTTGTATT--TCTAATGCTACTCAAAATGTTACTG 868  
 Db 670 TTTCTGACGCAATTCGACCGGCTCTTGTGCTGGAAGGAGGAAACCAACGCTTGTG 729  
 Qy 869 TTAAGCTTGAGATTACTCAAAATCATATTTTCAGCTCATGTTAACTCACTGAGGTTGTA 928  
 Db 730 AAATGACGACGCTTCTCGGAGGATGCTCGAAATTTCTCGATTCTCCGAGATCGGA 789  
 Qy 929 AGGTAAACGAGATTTAGATATTTGATGCTAGTCTGTTAGTGAATTAATCACTCA 988  
 Db 790 AAGTGACGAGCAGCTGCAACGAGTCTCTTGAAGGAGGAGTTGAACATTTCTGTTCA 849  
 Qy 989 ATCCAACTTACCTTATGAGTCTTTTAAAGCTATTAAAGTGAACAGTGAATAATCA 1046  
 Db 850 GCGGAAATATATGATGACGCTTGGCGGCTTGTGGAACAGACATTTCAAATCA 907

RESULT 33  
 US-10-673-098-173  
 ; Sequence 173, Application US/10673098  
 ; Publication No. US20040048309A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: O'Donnell, Michael E.  
 ; APPLICANT: Yuzhakov, Alexander  
 ; APPLICANT: Yurieva, Olga  
 ; APPLICANT: Jeruzalmi, David  
 ; APPLICANT: Bruck, Irina  
 ; APPLICANT: Kuriyan, John  
 ; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT  
 ; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND  
 ; TITLE OF INVENTION: USE THEREOF  
 ; FILE REFERENCE: 2221/1030  
 ; CURRENT APPLICATION NUMBER: US/10/673,098  
 ; CURRENT FILING DATE: 2003-09-25  
 ; PRIOR APPLICATION NUMBER: US/09/716,964A  
 ; PRIOR FILING DATE: 2000-11-21  
 ; PRIOR APPLICATION NUMBER: 60/143,202  
 ; PRIOR FILING DATE: 1997-04-08  
 ; PRIOR APPLICATION NUMBER: 08/823,407  
 ; PRIOR FILING DATE: 1997-04-08  
 ; PRIOR APPLICATION NUMBER: 09/057,416

;; PRIOR FILING DATE: 1998-04-08  
;; NUMBER OF SEQ ID NOS: 212  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 173  
;; LENGTH: 992  
;; TYPE: DNA  
;; ORGANISM: Bacillus stearothermophilus  
US-10-673-098-173

Query Match 7.8%; Score 88; DB 16; Length 992;  
Best Local Similarity 46.5%; Pred. No. 2.4e-08;  
Matches 390; Conservative 0; Mismatches 440; Indels 8; Gaps 3;  
Qy 214 CCAGGAGCTATTTTATAGAGCTAGTGTGTTTATTAATATATTTTCAAGTTTGCCAGAT 273  
Db |||||  
Qy 73 CCGGGAGCATCGTACTGCGAGCGCTTTTCTCTGAATCGTGAATAACTGCCGAA 132  
Db |||||  
Qy 274 ATTAGTATAATGTTAAAGAAATGAAACACCAAGTGTGTTTAAACAGTGGTAATCA 333  
Db |||||  
Qy 133 CAACCGTGAATCGAAACGGAAGACAACTTTTGACGATCATCGCTCGGGGCACTCA 192  
Qy 334 GAGATTACCTTAAAGGAAAGATGTTGACCATCTCTCGTCTACAAAGATCATCA 393  
Db |||||  
Qy 193 GAATTCGGCTCAATCGGCTAAACGCGGAGATATCCGCGCTCGCCAAATGAAGAA 252  
Qy 394 GAAATCCCTTTGATTTTAAACAAATTTATGAGTCTATTTGCTGAAACAGCTTTT 453  
Db |||||  
Qy 253 GAAACCGTGTTCCTGATTCGAGCGGCTGATTTTGAACCGTGTATTCGGCAACCGGTGTC 312  
Qy 454 GCAGCCAGTTCACAGATGATTTGAGACCGTGTGAGTATTTTCTCACCAGCA 690  
Db |||||  
Qy 490 CTCAGCAAAATTTTGGATGACGCGCAACACCGGCTGGACATCGTATGACGCAATCA 549  
Qy 691 ATCTTGTTCAGAGTGAACACATTTCTTTTATACACGCTCTTGAAGAGAAATATCC 750  
Db |||||  
Qy 550 GTGCTATTTAAGCGGAGCAGCTTCTCTCTTTTCCGCGCTGCTGACGCACTATCCG 609  
Qy 751 GATACAGACCGTTCATGACAGAAATTTGAGACGAGGTTGTTTCAATACCAATCC 810  
Db |||||  
Qy 610 GAGACGCGCGCTTGAATCCACAGAAAGCAACGACCATGATGTCATGCAATGCAAAAG 669  
Qy 811 CTTGCCACGCTATGGAACGCTCTCTTTGATTTTCTAATGCTACTCAAAATGCTACTG 868  
Db |||||  
Qy 670 TTTCTGAGGCAATCGACGAGCGCTTGTCTGTCGAGAGGAGAGCAACGTTG 729  
Qy 869 TTAAGCTTGAATTAATCAATCAATTTTCAAGTCTATGTTAACTCAGCTGAGTTGTA 928  
Db |||||  
Qy 730 AAATGACGAGCTTCTCGGAGGAATGCTGAAATTTTCTGATTTCTCGGATCGGA 789  
Qy 929 AGGTAAACGAGGATTTAGATTTGTTAGTCACTGCTGTTGATGTTTAACTATCAGTTCA 988  
Db |||||  
Qy 790 AAGTACGAGCAGCTGCAACGAGGCTCTTTGAAGGGGAGAGTTGAACATTTTCGTTCA 849  
Qy 989 ATCCAACTTACCTTATGAGTCTTTTAAAGCTATTTAAAGTGAACAGTAAATTC 1046  
Db |||||  
Qy 850 CGCGGAAATATATGATGACGCGTTGCGGCGCTTGTATGGAACAGACATTTCAATCA 907

RESULT 34  
US-10-672-638-173  
; Sequence 173, Application US/10672638

;; Publication No. US20040077012A1  
;; GENERAL INFORMATION:  
;; APPLICANT: O'Donnell, Michael E.  
;; APPLICANT: Yuzhakov, Alexander  
;; APPLICANT: Yuzhakov, Olga  
;; APPLICANT: Jeruzalmski, David  
;; APPLICANT: Bruck, Irina  
;; APPLICANT: Kuriyan, John  
;; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT  
;; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND  
;; FILE REFERENCE: 22221/1030  
;; CURRENT APPLICATION NUMBER: US/10/672,638  
;; CURRENT FILING DATE: 2003-10-09  
;; PRIOR APPLICATION NUMBER: US/09/716,964A  
;; PRIOR FILING DATE: 2000-11-21  
;; PRIOR APPLICATION NUMBER: 60/143,202  
;; PRIOR FILING DATE: 1997-04-08  
;; PRIOR APPLICATION NUMBER: 08/823,407  
;; PRIOR FILING DATE: 1997-04-08  
;; PRIOR APPLICATION NUMBER: 09/057,416  
;; PRIOR FILING DATE: 1998-04-08  
;; NUMBER OF SEQ ID NOS: 212  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 173  
;; LENGTH: 992  
;; TYPE: DNA  
;; ORGANISM: Bacillus stearothermophilus  
US-10-672-638-173

Query Match 7.8%; Score 88; DB 16; Length 992;  
Best Local Similarity 46.5%; Pred. No. 2.4e-08;  
Matches 390; Conservative 0; Mismatches 440; Indels 8; Gaps 3;  
Qy 214 CCAGGAGCTATTTTATAGAGCTAGTGTGTTTATTAATATATTTTCAAGTTTGCCAGAT 273  
Db |||||  
Qy 73 CCGGGAGCATCGTACTGCGAGCGCTTTTCTCTGAATCGTGAATAACTGCCGAA 132  
Db |||||  
Qy 274 ATTAGTATAATGTTAAAGAAATTTGACACCAAGTGTGTTTAAACAGTGGTAATCA 333  
Db |||||  
Qy 133 CAACCGTGAATTCGAAACGGAAGACAACTTTTGAACGATCATCGCTCGGGGCACTCA 192  
Qy 334 GAGATTACCTTAAAGGAAAGATGTTGACCATCTCTCGTCTACAAAGATCATCA 393  
Db |||||  
Qy 193 GAATTCGGCTCAATCGGCTAAACGCGGAGATATCCGCGCTCGCCAAATGAAGAA 252  
Qy 394 GAAATCCCTTTGATTTTAAACAAATTTATGAGTCTATTTGCTGAAACAGCTTTT 453  
Db |||||  
Qy 253 GAAACCGTGTTCCTGATTCGAGCGGCTGATTTTGAACCGTGTATTCGGCAACCGGTGTC 312  
Qy 454 GCAGCCAGTTCACAGATGATTTGAGACCGTGTGAGTATTTTCTCACCAGCA 513  
Db |||||  
Qy 313 GCGGTTTCTACATCGGAACGCGCCCAATCTTGACAGGTTCACTGGAAGAGT---TGAA 369  
Qy 514 CATAAAGATTTTAAAGCAGTACGCTCTCTCATCGTATGAGCAACGTTTAACTCACT 573  
Db |||||  
Qy 370 CATGGCGAGCTGCTCTGCAAGCGGACCGGATCATCGCTTAGCCATCGCAAGTGA 429  
Qy 574 TTGGA---ATACTTTCAGCAGATTTGATGTTAGTGTGTTTCAAGTGAATTTTGAAGAA 630  
Db |||||  
Qy 430 ATTGAGTCGGAATGAAGATTCATACACGCTGTCATCCCTGGAAAGAGTCTTATAG 489  
Qy 631 TTTTCAGCAGTATTTACAGATGATTTGAGACCGTGTGAGTATTTTCTCACCAGCA 690  
Db |||||  
Qy 490 CTCAGCAAAATTTTGGATGACGCGCAACACCGGCTGGACATCGTATGACGCAATCA 549  
Qy 691 ATCTTGTTCAGAGTGAACACATTTCTTTTATACAGCTCTTGAAGAGAAATTTATCC 750  
Db |||||  
Qy 550 GTGCTATTTAAGCGGAGCAGCTTCTCTCTTTTCCGCGCTGCTGACGCACTATCCG 609  
Qy 751 GATACAGACCGTTCATGAGTATTTGAGACCGGAGTGTGTTTCAATACCAATCC 810  
Db |||||  
Qy 610 GAGACGCGCGCTTGAATCCACAGAAAGCAACGACCATGATGTCATGCAATGCAAAAG 669

QY 811 CTTCGCCAGCTATGGAACGGTGCCTTCTTGATT--TCTAATGCTACTCAAAATGGTACTG 868  
 DB 670 TTCTGCGAGCAATCGACCGAGCGCTTGTCTGCTCGAGAAGGAAGCAACAACTGCTG 729  
 QY 869 TTAAGCTTGAAGTAACTCAAAATCATATTTTCAGCTCATGTTAACTCACTGAGGTTGGTA 928  
 DB 730 AAATGACGAGCGCTTCTGAGGAATGCTCGAAATTTCTCGAATTTCTCCGAGATCGGGA 789  
 QY 929 AGGTAACGAGGATTTAGATATTTGTAGTCACTGCTGCTGAGTGAATTAATACTATCAGCTTCA 988  
 DB 790 AAGTGACGAGCAGCTGCAACCGAGTCTCTTGAAGGGGAAGAGTTGAACATTTGTTCA 849  
 QY 989 ATCCAACCTTACCTTATGAGCTTTTAAAGCTATTTAAAGTGAAACAGTAAAAATCA 1046  
 DB 850 GCGGAAATATATGATGAGCGGCTTGGCGGCGCTTGATGGAACAGACATTTCAATCA 907

RESULT 35  
 US-10-673-127-173  
 ; Sequence 173, Application US/10673127  
 ; Publication No. US20040081995A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: O'Donnell, Michael E.  
 ; APPLICANT: Yuzhakov, Alexander  
 ; APPLICANT: Yurieva, Olga  
 ; APPLICANT: Jeruzalmi, David  
 ; APPLICANT: Bruck, Irina  
 ; APPLICANT: Kuriyan, John  
 ; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT  
 ; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND  
 ; TITLE OF INVENTION: USE THEREOF  
 ; FILE REFERENCE: 2221/1030  
 ; CURRENT APPLICATION NUMBER: US/10/673,127  
 ; PRIOR FILING DATE: 2003-09-26  
 ; PRIOR FILING DATE: 2000-11-21  
 ; PRIOR APPLICATION NUMBER: 60/143,202  
 ; PRIOR FILING DATE: 1997-04-08  
 ; PRIOR APPLICATION NUMBER: 08/823,407  
 ; PRIOR FILING DATE: 1997-04-08  
 ; PRIOR APPLICATION NUMBER: 09/057,416  
 ; PRIOR FILING DATE: 1998-04-08  
 ; NUMBER OF SEQ ID NOS: 212  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 173  
 ; TYPE: DNA  
 ; ORGANISM: Bacillus stearothermophilus

US-10-673-127-173  
 Query Match 7.8%; Score 88; DB 16; Length 992;  
 Best Local Similarity 46.5%; Pred. No. 2.4e-08;  
 Matches 390; Conservative 0; Mismatches 440; Indels 8; Gaps 3;

QY 214 CCAGAGCTATTTTATGAGCTAGTTTTTTTATTAATATATTTTCAAGTTTGCAGAT 273  
 DB 73 CCGGAGGAGCATGTACTGCGGCGCTTTTCTCTGAATCGTGAATAAAGTCCGCA 132  
 QY 274 ATTAGTATAATGTTAAAGAAATTAACAACCAAGTGTGTTTAAACAGTGTAAATCA 333  
 DB 133 CAACCGGTGGAATCGAACCGGACGATATTTTGGAGCATCATCCGCTCGGGGCACTCA 192  
 QY 334 GAGATTACCTTAAAAAGGAAAGATGTTGACCAAGTATCTCTGCTTCAAGAAGTATCAACA 393  
 DB 193 GAATTCGGCTCAATGGGCTAAACCGGACGATATCGGCGCTCGCGCAATTAAGAA 252  
 QY 394 GAAATCTCTGATTTTAAAAACAAATTTATGAGTCTATTTATGCTGGAACAGCTTTT 453  
 DB 253 GAAACGCTGTTTCAAAATCCGGCTGATTTATGAAACCGTGTATTCGGCAACGCTGTC 312  
 QY 454 GCAGCCAGTTTCAAGAAAGTCTCTATTTTAAACAGGATTCATTTGTTATTAAGTAA 513

DB 313 GCCGTTTCTACATCGGAAACCGGCCCAATCTTGACAGGTGTCAACTGGAAAGT---TGAA 369  
 QY 514 CATAAAGATTTTAAAGCAGTAGCAGCTGACTCTCATCGTATGAGCCAACTGTTTAACTACT 573  
 DB 370 CATCGGAGCTTGTCTGCACAGCGACGACAGTATCGCTTTAGCCATCGCCAAAGTGAAA 429  
 QY 574 TTGACA---ATACTTCAGCAGATTTGATGTAGTCTTCTCAAGTAAATCTTTGAGAGAA 630  
 DB 430 ATTGAGTCGGAAATGAAGTATCATCAACGTCGTCATCCCTGGAAAAAGTCTTTAATGAG 489  
 QY 631 TTTTCAGCAGTATTTACAGATGATATTGAGACCGTTGAGGTATTTTCTCACCAGCCAA 690  
 DB 490 CTCAGCAAAATTTTGGATGACGCAACCAACCGGTGACATCGTCATGACAGCCCAATCA 549  
 QY 691 ATCTGTTTCAAGTGAACACATTTCTTTTATACAGCTCTTTAGAGGAAATTAATCC 750  
 DB 550 GTGCTATTAAAGCGGAGCACTTCTCTCTTTTCCCGCTGCTTGACGGCAACTATCCG 609  
 QY 751 GATACAGACCGTTTAAATGACAGAAATTTGAGACGGAGGTGTTTCAATACCCCAATCC 810  
 DB 610 GAGACGGCCCGCTTGATTCCAAACAGAAAGCAACGACCATGATCGTCAATGCAAAAGAG 669  
 QY 811 CTTCGCCAGCTATGGAACGTCCTTCTTGATT--TCTAATGCTACTCAAAATGGTACTG 868  
 DB 670 TTCTGCGAGCAATCGACCGAGCGCTTGTCTGCTCGAGAAGGAAGCAACGTTGTG 729  
 QY 869 TTAAGCTTGAATTAATCAAAATCATATTTTCAGCTCATGTTAACTCACTGAGGTTGGTA 928  
 DB 730 AAATGACGAGCTTCTGAGGAATGCTCGAAATTTCTCGAATTTCTCCGAGATCGGGA 789  
 QY 929 AGGTAACGAGGATTTAGATATTTGTAGTCACTGCTGCTGAGTGAATTAATACTATCAGCTTCA 988  
 DB 790 AAGTGACGAGCAGCTGCAACCGAGTCTCTTGAAGGGGAAGAGTTGAACATTTCTGTTCA 849  
 QY 989 ATCCAACCTTACCTTATGAGCTTTTAAAGCTATTTAAAGTGAAACAGTAAAAATCA 1046  
 DB 850 GCGGAAATATATGATGAGCGGCTTGGCGGCGCTTGATGGAACAGACATTTCAATCA 907

RESULT 36  
 US-10-670-817-173  
 ; Sequence 173, Application US/10670817  
 ; Publication No. US20040106137A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: O'Donnell, Michael E.  
 ; APPLICANT: Yuzhakov, Alexander  
 ; APPLICANT: Yurieva, Olga  
 ; APPLICANT: Jeruzalmi, David  
 ; APPLICANT: Bruck, Irina  
 ; APPLICANT: Kuriyan, John  
 ; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT  
 ; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND  
 ; TITLE OF INVENTION: USE THEREOF  
 ; FILE REFERENCE: 2221/1030  
 ; CURRENT APPLICATION NUMBER: US/10/670,817  
 ; CURRENT FILING DATE: 2003-09-25  
 ; PRIOR APPLICATION NUMBER: US/09/716,964  
 ; PRIOR FILING DATE: 2000-11-21  
 ; PRIOR APPLICATION NUMBER: 60/143,202  
 ; PRIOR FILING DATE: 1997-04-08  
 ; PRIOR APPLICATION NUMBER: 08/823,407  
 ; PRIOR FILING DATE: 1997-04-08  
 ; PRIOR APPLICATION NUMBER: 09/057,416  
 ; PRIOR FILING DATE: 1998-04-08  
 ; NUMBER OF SEQ ID NOS: 212  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 173  
 ; LENGTH: 992  
 ; TYPE: DNA  
 ; ORGANISM: Bacillus stearothermophilus

US-10-670-817-173  
 Query Match 7.8%; Score 88; DB 17; Length 992;



Db 790 AAGTGACGGAGCAGCTGCAACCGAGTCTTTGAAGGGGAAGAGTTGAACATTTTCGTCA 849  
Qy 989 ATCCAACTTACCTTATGAGTCTTTAAAGAGCTATTAAAGAGTGAACAGAGTAAATTC 1046  
Db 850 GCGCGAATATATGATGAGCGGTTGCGGGCGCTTGATGGAACAGACATTTCAATCA 907

RESULT 38  
US-10-671-207-173  
; Sequence 173, Application US/10671207  
; Publication No. US20040197796A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Donnell, Michael E.  
; APPLICANT: Yuzhakov, Alexander  
; APPLICANT: Yurieva, Olga  
; APPLICANT: Jeruzalmski, David  
; APPLICANT: Bruck, Irina  
; APPLICANT: Kuriyan, John  
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT  
; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND  
; FILE REFERENCE: 22221/1030  
; CURRENT APPLICATION NUMBER: US/10/671,207  
; CURRENT FILING DATE: 2003-09-25  
; PRIOR APPLICATION NUMBER: US/09/716,964  
; PRIOR FILING DATE: 2000-11-21  
; PRIOR APPLICATION NUMBER: 60/143,202  
; PRIOR FILING DATE: 1997-04-08  
; PRIOR APPLICATION NUMBER: 08/823,407  
; PRIOR FILING DATE: 1997-04-08  
; PRIOR APPLICATION NUMBER: 09/057,416  
; PRIOR FILING DATE: 1998-04-08  
; NUMBER OF SEQ ID NOS: 212  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 173  
; LENGTH: 992  
; TYPE: DNA  
; ORGANISM: Bacillus stearothermophilus  
US-10-671-207-173

Query Match 7.8%; Score 88; DB 18; Length 992;  
Best Local Similarity 46.5%; Pred. No. 2.4e-08;  
Matches 390; Conservative 0; Mismatches 440; Indels 8; Gaps 3;

Qy 214 CCAGGAGCTATTTTATGAGCTAGTCTTTTATTAATATTTTCAAGTTTGCAGAT 273  
Db 73 CCGGGGAGCATCTGCTGAGCGCGCTTTTCTCTGAATCTGAAAACTGCGCA 132  
Qy 274 ATTAGTAAATGTTAAAGAAATTGAACAAACCAAGTTGTTTAAACAGTGGTAAATCA 333  
Db 133 CAAACGGTGGAAATCGAAACGGAAGCAACTTTTGAAGATCATCGCTCGGGGCACTCA 192  
Qy 334 GAGATTACCTTAAAGGAAAGATGTTGACAGTATCTCTGCTACAGAGTATCAACA 393  
Db 193 GAATTCGGCTCAATGGGGCTAAACCGGAGCAATATCGCGCTCGCGCAATTTGAAGAA 252  
Qy 394 GAAATTCCTTTCATTTTAAACAAATTTATGAGTCTTATTTGCTGAAACAGCTTTT 453  
Db 253 GAAACGTTGTTTCAATCCCGCTGATTTTGAACCGTGTATTCGCAACCGTGTTC 312  
Qy 454 GAGCGAGTTTCAAGAAAGTCTGCTCTATTTTAAAGAGTTTCAATTTGTTATTAAGTAT 513  
Db 313 GCGGTTTCTACATCGGAAACGCGGCAATCTTGACAGGTGTCAACTGGAAAGT---TGAA 369  
Qy 514 CATAAGTTTAAAGCAGTACGCTGCTCTCTGCTATGCTATGAGCAACAGTTTATCACT 573  
Db 370 CATGGGAGCTGTCTGCAACGCGACCGGAGTCTGCTTAGCCATGCGCAAGTGA 429  
Qy 574 TTGGACA---ATACCTTACAGAGTTTGTAGTGTCTTCCAAAGTAAATCTTTTGAGAA 630  
Db 430 ATTGAGTCGGAATGAAGTATCATACACGCTGCTCATCCCTGGGAAAGTCTTATGAG 489

Qy 631 TTTTCAGCAGTATTTTACAGATGATTTGAGACCGTTGAGGTATTTTCTCACCAGCCAA 690  
Db 490 CTCAGCAAAATTTTGGATGACGGCAACCAACCGGTGGACATCGTCATGACAGCCCAATCAA 549  
Qy 691 ATCTTTTTCAGAGTGAACACATTTCTTTTATACACGCTCTTTAGAGGAAATTTATCCC 750  
Db 550 GTGCTATTTAAGCGGAGCACCTTCTCTTTTCCCGGCTGCTTGACGGCACTATCCG 609  
Qy 751 GATACAGACCGTTTATTAATGACAGAAATTTGAGACGGAGGTTGTTTTCAATAACCAATCC 810  
Db 610 GAGACGGCCCGCTTATTTCCAAACAGAAAGCAACCAATGATCGTCAATGCAAAAGAG 669  
Qy 811 CTTGCCACGCTATGGAACGTCCTTCTTGATT--TCTAATGCTACTCAAAATGCTACTG 868  
Db 670 TTTCTGACGGAATCGACCGGCTCTTGTCTGCGAAGGAGGAAACACGCTTGTG 729  
Qy 869 TTAAGCTTGTAGTACTCAAAATCATATTTTCAGCTCATGTTAACTCACCTGAGGTTGGTA 928  
Db 730 AAACGTGACGCGCTTCTGAGGAGATGCTCGAAATTTCTTCGATTTCTCCGAGATCGGA 789  
Qy 929 AGGTAAACGAGGATTTAGATATTTGTAGTCACTGCTGCTGAGTATTAATACTATCAGCTTCA 988  
Db 790 AAGTGACGAGCAGCTGCAACGAGTCTCTTGAAGGGGAAGAGTTGAACATTTTCGTTCA 849  
Qy 989 ATCCAACTTACCTTATGAGTCTTTTAAAGCTATTTAAAGTGAACAGTAAAAATTC 1046  
Db 850 GCGGAAATATATGATGAGCGCTTTCGCGGCGCTTGTGGAACAGACATTTCAAAATCA 907

RESULT 39  
US-10-282-122A-15836  
; Sequence 15836, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA 034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1



```

; SEQ ID NO 15836
; LENGTH: 1101
; TYPE: DNA
; ORGANISM: Clostridium
US-10-282-122A-15836

```

Query Match 6.5%; Score 74; DB 16; Length 1101;  
Best Local Similarity 46.0%; Pred. NO. 2.1e-05;  
Matches 503; Conservative 0; Mismatches 555; Indels 36; Gaps 6;

QY 4 ATTCAAATTTTCAATTAATCGCAACATTAATTAATTCATGCTTTAAATACAACCTAAAGCTGCT 63  
DB |||||  
1 ATGAAAATTATATGTGTAAGAAAATATATATACAGAGCAATTAATTAACAGCTCAAAAAGCT 60  
QY 64 ATTAGCACTAAAATGCGCAATTCCTATCTTTTATCAATAAAAAATTTGAAGTCACTTCTACA 123  
DB |||||  
61 GTTACAGAAAAATCAACAATGCCAGTATTTACAAGGTATACTCATGAGTGTTCAAAATAAC 120  
QY 124 GGAGTAACCTTTAAACAGGGTCTAACGGTCAAAATATCAATTTGAAACACTATTTCTGTGAAGT 183  
DB |||||  
121 GAATTAACCTTTAATTTGGTCTGATATAGATTTAAGTATGAAACTTAAATAAATGTAGAA 180  
QY 184 AATGAAAAATGCTGGTTTGTCTAATTAACCTCTCCAGGACTATTTATTAGAGCTAGTTTTT 243  
DB |||||  
181 GTTTTGAAGAAGGTAAGTTGTT-----TTAGATGCTAGGCTT 219  
QY 244 TTATTATAATTAATTTCAAGTTTGCAGATATTAGTATAAATGTTAAAGAAATTTGAACAA 303  
DB |||||  
220 TTAAGTGAATTAATTAGAAAGCTGCCTAATTTCCAAGGTTGAGATACAAACTATTGAGAA 279  
QY 304 CACCAAGTTGTTTTAAACAGTGGTAAATCAGAGATTACCTTAAAGGAAAGATGTTGAC 363  
DB |||||  
280 AATGTGTTTGAATAAATCTGTATTAATCAAGCTAACTGTGTATCTTAAATCCAAAT 339  
QY 364 CAGTATCCTCTGTACAGAAGTATCAACAGAAAATCCTTTGATTTTAAAAACAAATAA 423  
DB |||||  
340 GAATTTCCAAGCTTACCAGAGATAGAGAAAATCTCAATCTTCAAAATAATCAAAAACA 399  
QY 424 TTGAAGTCTATTAATGCTGAAACAGCTTTTTCAGCCAGTTTACAGAAAGTCGTCTATT 483  
DB |||||  
400 TTAAGAACTATGATAAAGGTACTATTAATTTGCTATTGCTCAGAGTAAAAAAGGCTATA 459  
QY 484 TTAACAGAGGTTCAATATGTTATTAAAGTAATCATATAAGATTTTAAAGCAGTAGCGACTGAC 543  
DB |||||  
460 TTAACAGCGCTTTTATTGAAATA--AAAGATAGCAATTAATTTAGTCTGCTATTGAT 516  
QY 544 TCTCATCTATGAGCCCAAGCTTTTAATCACTTTGGCAATATCTTCCAGAGATTTGATGTA 603  
DB |||||  
517 GGATATAGGCTAGCTTTAAG---AAGCAATACATAGATAAGCAACCGATATAAATGCA 573  
QY 604 GTTCTTCCAAGTAAATCTTTGAGAGAATTTTCAGCAGTATTTACAGATGATATTGAGACC 663  
DB |||||  
574 GTAATTCAGGAAAAACACTTAATGAAGTTTAAATTTTGAAGATGATGTTGA---T 630  
QY 664 GTTGAGGTATTTTCTCAACAGCCAAATCTTTGTTCAAGAGTGAACACATTTCTTTTAT 723  
DB |||||  
631 GTAAATATTACATTTTACTTCTACCATATAATTATTCAATCTTTGGAATTAAGTAATA 690  
QY 724 ACAGCCCTCTTTAGAAAGAAATATCCGATACAGACCGTTTATTATGACAGAATTTGAG 783  
DB |||||  
691 TCAAGACTACTTGAAGGTGAATTTATAAATAAATTAATCAATAAATACCTTGAAGAATAAAC 750  
QY 784 ACGAGGTGTTTTTCAATACCAATCCCTTCGCCACGCTATGGAAGGTCCTTCTTGATT 843  
DB |||||  
751 TTAATATTGTGCTAGAAAAGAAATTTGTTAGACTCATAGAGAGCTTCACTTATG 810  
QY 844 TCTAATGCTACTCAAAATGGTACTGTTAAGCTTTGAGATTTACTCAAAATCATATTTAGCT 903  
DB |||||  
811 GCTAAGATGGAATAACAAATTTGATAAATTTAGACAT---CGAGGATGATGTAATGAT 867  
QY 904 CATGTTTAACCTACCTGAGGTTGGTAAGGTAACAGGATTTAGATATTGTTAGTCACTGCT 963  
DB |||||  
868 ATAAAGTCCAAATTCATTTGGAAACGTAGAGAAGAAATAAATTTTGA---A 924

Qy 964 GGTAGTGAATTAATCATCACTCAAGCTTCAATCCAACTTACCTTATGTAGTCTTTAAAAAGCTATT 1023  
Db 925 GGTGAGCCTTTGAAATTTGCAATTTAACTCAAAATATTTAATAGATGCTTAAAAATCATG 984  
Qy 1024 AAAAGTGAACACAGTAAAAAATTCATTTCTATCACCAGTTCCAGCACTTCACTTAAACACCA 1083  
Db 985 AATCAGGAAGAAATTTGTAATGAACTTTTCAAGTAGCATAAGTCTTGCATTTATTTAAAAAT 1044  
Qy 1084 GGCGATGAGGAAGA 1097  
Db 1045 AAGGAAAATGATGA 1058

```

RESULT 40
US-09-974-300-6041
; Sequence 6041, Application US/09974300
; Patent NO. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berkla, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6041
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Bacillus clausii
US-09-974-300-6041

```

Query Match 6.0%; Score 68.6; DB 9; Length 486;  
Best Local Similarity 48.8%;  
Matches 216; Conservative 0; Mismatches 224; Indels 3; Gaps 1

QY	664	GTTCAGGTATTTTCTTCACCAAGCCAAATCTTGTTCAGAAGTGAACACATTCTCTTTTAT	723
Db	15	GCTGAAATTGTTGTAACCGAGGAATCAAACTCTGTTCAACTCGGAACATTGTTGTTTTC	74
QY	724	ACACGGCTCTTAGAGGAAATATCCCGATACAGACCGTTTATTAATGACAGAATTGAG	783
Db	75	TCAGAGCTCTTGAAGGGAAGTATCCAGTAACGTCGAGCATGATACCGACTCAGTTCAA	134
QY	784	ACGGAGGTTGTTTTCATATACCAATCCCTTCGCCACGCTATGAACGTCGCTTCGTGATT	843
Db	135	ACGTCCTTTGTCGTAAAAACAAGAATGTCGTCATACGCTTGAACGGGCAATGTTGTG	194
QY	844	TCTAATGCTACTCAAAATGTGTTAAGCTTTGAGA--TTACTCAAAATCATATTTCA	900
Db	195	TCAAGGGAAGCAAAAACAATGTCAATTAACGTCAAAACACTTTGGCGACAACCAATTGAA	254
QY	901	GTCATGTTAACTCACTGAGCTGAGTTGTGAAGGTAAACGAGGATTAGATATTGTTAGTCAG	960
Db	255	ATTACGTCACATCAACAAGAATAGGGAAGTACCGAANAATGCCACTGACCAATTT	314
QY	961	TCGTGTAGTGAATTAACCTATACGCTTCAATCCAACTTACCTTTATGAGTCTTTAAAAGCT	1020
Db	315	GAAGGGGATGAAATCCGGANTTCGTTTAATGGGAAAATATTCCTTGATGCCCTGAAAGTG	374
QY	1021	ATTAAAAGTGAACAGTAGAAAATTCATTTCTATACAGTTGCAACATTACCCCTAACA	1080
Db	375	GTTGACACGATGAAGTCCAAATCTTTTTGACAGGTGCGATGAGTCCCTTCGTTATTCCG	434
QY	1081	CCAGCGGATGAGGAAGAAGTTT	1103
Db	435	CCAAATTGATTCAGACCACTATTT	457

Fri Feb 4 10:26:48 2005

us-10-048-071-27.rnpb

Page 33

Search completed: January 31, 2005, 13:50:53  
Job time : 508 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 16:54:57 ; Search time 40 Seconds  
(without alignments)  
626.705 Million cell updates/sec

Title: US-10-048-071-28  
Perfect score: 1863  
Sequence: 1 MIQFSINRTLFTHALNTTKRAISTKNAIPILSSIKIEVTSTGTGTLTSGNQGSIENFTIPV 378

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A-COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B-COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A-COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B-COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCITUS-COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1389	74.6	378	4	US-09-583-110-4405
2	863.5	46.3	384	4	US-09-107-532A-4556
3	748.5	40.2	385	3	US-09-134-001C-5166
4	431	23.1	181	4	US-09-134-000C-5039
5	350.5	18.8	368	4	US-09-252-991A-18904
6	315	16.9	383	4	US-09-328-352-5969
7	313	16.8	374	4	US-09-489-039A-10579
8	307	16.5	375	4	US-09-543-681A-7400
9	249.5	13.4	375	4	US-09-818-780-23
10	232	12.5	303	4	US-09-198-452A-357
11	131.5	7.1	141	4	US-09-134-000C-5040
12	117.5	6.3	977	4	US-09-248-796A-15579
13	115.5	6.2	800	3	US-08-776-265-3
14	115.5	6.2	800	4	US-09-398-184-3
15	114.5	6.1	823	4	US-09-248-796A-16699
16	113.5	6.0	922	4	US-09-883-134-9
17	112.5	6.0	470	4	US-09-248-796A-23131
18	112.5	6.0	619	3	US-09-066-046-2
19	109.5	5.9	1753	4	US-09-248-796A-19154
20	109	5.9	532	4	US-09-710-279-546
21	107.5	5.8	569	4	US-09-248-796A-16697
22	107	5.7	567	3	US-09-134-001C-3172
23	106.5	5.7	1155	4	US-09-543-681A-6286
24	104.5	5.6	1726	4	US-09-700-227-2
25	103.5	5.6	564	4	US-09-107-532A-5248
26	103.5	5.6	920	4	US-09-463-402-6
27	103.5	5.6	921	4	US-09-889-572-4

ALIGNMENTS

RESULT 1

US-09-583-110-4405  
; Sequence 4405, Application US/09583110  
; Patent No. 6699703  
; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; FILE REFERENCE: Pneumoniae for Diagnostics and Therapeutics

; CURRENT APPLICATION NUMBER: US/09583,110  
; CURRENT FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/107,433

; PRIOR FILING DATE: 1998-06-30

; PRIOR APPLICATION NUMBER: US 60/085,131

; PRIOR FILING DATE: 1998-05-12

; PRIOR APPLICATION NUMBER: US 60/051,553

; PRIOR FILING DATE: 1997-07-02

; NUMBER OF SEQ ID NOS: 5322

; SEQ ID NO 4405

; LENGTH: 378

; TYPE: PRT

; ORGANISM: Streptococcus pneumoniae

US-09-583-110-4405

Query Match 74.6%; Score 1389; DB 4; Length 378;  
Best Local Similarity 72.2%; Pred. No. 6.7e-125; Mismatches 55; Indels 0; Gaps 0;  
Matches 273; Conservative 50;

QY	1	MIQFSINRTLFTHALNTTKRAISTKNAIPILSSIKIEVTSTGTGTLTSGNQGSIENFTIPV	60
DB	1	MHFSINKNLFQALNTTKRAISSKNAIPILSTKIDVTNEGITLGSNGQSIENFISQ	60
QY	61	SNENAGLITSPGAILLEASFPINISLDPISINVKIEHQVLTTSKSEITLKGQV	120
DB	61	KNEADGLITSLGSLLEASFPINVSLLPVDVLTKEIQNQIVLTSGSEITLKGDS	120
QY	121	DOYPLQVSTENPLILKTLKLSIIAETAFASIQESRPILTVGHVILSNHKDFKAVAT	180
DB	121	BOYPRIQISASTPILLETKLKKIINETAFASIQESRPILTVGHVILSNHKELKTAV	180
QY	181	DSHRMSQRILITDNTSADLMVLPSPKSLREFSAVFTDDIETVEVFPSPQILFRSEHSIF	240
DB	181	DSHRLSQKLTLEKNSDDFDVVPSPSLREFSAVFTDDIETVEIFFANQILFRSEHSIF	240
QY	241	YTRLLEGNYPTDRLMLTTEFEVVFNTQSLRHAMERAPLISNATONGTVKLEITQNHIS	300
DB	241	YTRLLEGNYPTDRLIPTDFNTTITFNVNLRQSMERARLLSSATQNGTVKLEIKDGVS	300
QY	301	AHVNSPEVKVNEDDLIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFSPVRPTLT	360
DB	301	HHVNSPEVKVNEDDLIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFSPVRPTLT	360

Db 301 AHVHSPVGVKNEEDTDQVTGDLTISFNPTYLIDSLKALNSEKVTISFISAVRPFTLV 360  
Qy 361 PGDEEESFIQLITPVRTN 378  
Db 361 PADTDEDFMQLITPVRTN 378

RESULT 2  
US-09-107-532A-4556  
; Sequence 4556, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 4556:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 384 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...384  
; SEQUENCE DESCRIPTION: SEQ ID NO: 4556:  
US-09-107-532A-4556  
Query Match 46.3%; Score 863.5; DB 4; Length 384;  
Best Local Similarity 46.6%; Pred. No. 2.1e-74;  
Matches 176; Conservative 72; Mismatches 129; Indels 1; Gaps 1;  
Qy 1 MIQFSINRFLTHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTSGNGQISIENTIPV 60  
Db 8 INKVLNRAFQWELQTVQRAISSKTTIPILGVKLTITQEGTLTGSNADISIEITLSV 67  
Qy 61 SNEAGLLITSPGAILLESFPIINISLPDISINVKIEQHVLTSGKSIITLKGKDV 120  
Db 68 ENEKANNQIESTGSIVLQARFFSEIIRLPETFTILEVLENQVAITSGKANFVINGLDA 127  
Qy 121 DOYPRLOEVSTENPLILKTLKLSITAEAFASLQESRPILTGVHIVLSNKHDKFVAT 180  
Db 128 DNYPHLPVNVESHQMKLPVHVLTKLINETVFAVSQHESRPILTGVHIFILSDN-SLLAVAT 186

Qy 181 DSHRMSORLITLONTADSADLMVLPKSLRBFSAVFTDDIETVEVFFSPSQILFRSEHISF 240  
Db 187 DSHRLSORVIPVEQAAADHFDIVIPGKSLIELSLTNEEEIVEISIMENQVLFTETWYF 246  
Qy 241 YTRLLEGNYPDTRLMTETETEVVFNQTSRLRHAMERAFILSNATONGTVKLEITQNHIS 300  
Db 247 YSRLEGNYPDTRNLIPSSFNTEVEFSPFLAAIERASLLISHEGRNNIVRLSIRPDAVV 306  
Qy 301 AHVHSPVGVKNEEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHLHSPVRPFTLT 360  
Db 307 LYGNSPRIGKVEESLYTASSGDLDISFNDPKAALRAFQDMSIKVKFISAIRPFTLE 366  
Qy 361 PGDEEESFIQLITPVRTN 378  
Db 367 PTEGQVFIQLITPVRTN 384

RESULT 3  
US-09-134-001C-5166  
; Sequence 5166, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5166  
; LENGTH: 385  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5166

Query Match 40.2%; Score 748.5; DB 3; Length 385;  
Best Local Similarity 40.5%; Pred. No. 2.4e-63;  
Matches 153; Conservative 86; Mismatches 136; Indels 3; Gaps 3;  
Qy 1 MIQFSINRFLTHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTSGNGQISIENTIPV 60  
Db 9 MMEFTIKEDYFINGQLNDTLKAISSPTTLPILTGIKIDAKENEVILTSGDSSEISIEITIPK 68  
Qy 61 SNEAGLL-ITSPGAILLEASFFINISLPDISINVKIEQHVLTSGKSEITLKGKD 119  
Db 69 QVDGEEIVEITETGTVLPGRFFVDIIKKLPKKEVKLSTNEQFQTLITSGHSEFNLSGLD 128  
Qy 120 VDOYPRLOEVSTENPLILKTLKLSITAEAFASLQESRPILTGVHIVLSNKHDKFVAT 179  
Db 129 PDQVPLLEVRSDDAIQLSVKLNIAQTAFVNSTETRPVLTVGNWLIQDN-ELICTA 187  
Qy 180 TDSHRMSORLITLONTADSADLMVLPKSLRBFSAVFTDDIETVEVFFSPSQILFRSEHIS 239  
Db 188 TDSHRLAVRKLQLEDESENKVNIIIPGKALSINKIMSDSDIDIFFASNOVLPRVGNIN 247  
Qy 240 FYTRLLEGNYPDTRLMTETETEVVFNQTSRLRHAMERAFILSNATONGTVKLEITQNHIS 299  
Db 248 FISRLLEGHYPDTRLFPENYEIKLGINNGDFYHAIIDRASLLAREGGNNVKLSTGNELV 307  
Qy 300 SAHVNSPVGKVNEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHLHSPVRPFTLT 359  
Db 308 ELSSTSPSIGTVKEVNANDVEGGNKLISFNSKYMMDALKDAIDNDEVEFEGTWMKPFTL 367  
Qy 360 PGDEEESFIQLITPVRT 377  
Db 368 KPQD-DDSVTQLILPIRT 384

Db  
351 VHEADNDPSAYVVM<sup>1</sup>PMR 367

```

RESULTS 6
US-09-328-352-5969
; Sequence 5969, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOS
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5969
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5969

```

```

Query Match      16.9%; Score 315; DB 4; Length 383;
Best Local Similarity 24.3%; Pred. No. 1e-21;
Matches 95; Conservative 100; Mismatches 172; Indels 24; Gaps 12

QY      1 MIQPSINRTLFIHALNTTKAISTKPAIPILSSIKITEVTSTGVLTSNGSQISIENTIPV 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1 IVRLKIAKESLLNVLSHWGVAVERHTNLINSVKIQTNAQALITIGSDLEVELVASTAL 60

QY      61 SNNAGLLITSPGAILLEASFFINISSLPDIS - INVKEIEQHVVLTGSKSEITIKGKD 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      61 S-EGACL---BAGETTPARKLMEICKSLPTAALIDQITEDQRCILKSGNSRFVLGTLP 116

QY      120 VDQVPELQEVSTENPLIKT - KLLKSIITAAFAALSQESRPILTGVHIVLSNHHKDFKAV 178
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      117 AEDPPLLTTENSQGVQVQVQRELKRLFEKTAFAMAQVDVRFYLTGTLLEIDENQ-LRAV 175

QY      179 ATDSHRMSQRILITDNTSADLM - VLPFSKLSRFSFAVTTDDIETVEVFF-----S 227
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      176 TTDGHRLLCALCEIASSTSSQLQVAIVPRKAVGELQRLLSIEDQLTLLIGRELLNVTTNT 235

QY      228 PSQIILFRSE-HISFVTRILEGNYPDPTRLAMTREFTEVVENTOSLRHAMERAFLLSNATO 286

```

236	PSRQEQGGDITVRRETKLLIDGKFFDYRRVYPRGGDKHVLIGHDVFKQSQRVAIINSEKL	295
287	NGTVKLEITONHISAHVNSPEVGKVNBDLDIVSQSGDLTISFNPTYLIESLKAIKSETV	346
296	RG-VFLFNQDSLOQLRANNPQDEAIEDLAIQQS-APLEMSFNAQYLLDVLGVLGGDDV	353
347	KIHFLSPVRPFTL-TPGDESESFQLITPVR	376
354	NMSMTEAROSVLVDPAHPDQTVV--VMPNR	382





Db 232 LALALEGGGVRMALRLMEGEPDQVRVLPQFALRVQVEGALREAVRVSLSD-RQN 290  
QY 288 GTVKLEITQNHISAHVNSPEVGVKVNEDLDIVSQSGDLTISFNPTYLIESLKAISKSETVK 347  
Db 291 HRVDLLLEGRILLSAEG-DYCKGQEVF-AQVEGPGMAVAYNARVLLLEALAPV-CDRAH 347  
QY 348 IHFLSPVRFTTTPGDEESFQLITPVR 376  
Db 348 LGISGTSP-SLIWGD-GEGRVAVVPLR 374

RESULT 10  
US-09-198-452A-357  
; Sequence 357, Application US/09198452A  
; Patent No. 6559294  
; GENERAL INFORMATION:  
; APPLICANT: Griflais, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection  
; FILE REFERENCE: 9710-003-999  
; CURRENT APPLICATION NUMBER: US/09/198,452A  
; NUMBER OF SEQ ID NOS: 6849  
; SEQ ID NO 357  
; LENGTH: 303  
; TYPE: PRT  
; ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-357

Query Match 12.5%; Score 232; DB 4; Length 303;  
Best Local Similarity 25.4%; Pred. No. 6.6e-14;  
Matches 67; Conservative 63; Mismatches 124; Indels 10; Gaps 7;

QY 115 LKGVQDQVPRLOEVSTENPLILTKLKSIIAETAFASLOESRPILTGHVILSNKD 174  
Db 44 LLSMEKEDFPMLPDQNALRSLPABQLMORTSFVRSRESRVLTGVLALANGV- 102

QY 175 FKAVATDSHRMSQ--RLITDNTSADLMVLPKSLRFSFSAVFTDDIETVEVFSPSQIL 232  
Db 103 ATIVGTDGRLAKIDAEVTLDRSFG-EYIIPKAVEEIKMCSDEGEAT-IFLQODKIA 160

QY 233 FSEHISFTRLEGNYPDTDLMTETEVFVNTQSLRHAMERAFILSNATQNGTVKL 292  
Db 161 VECNDTLITLKLSSGEPDFSPVISTESNVKLDLHREELITLLKQVALFTNESSH-SVKF 219

QY 293 EITQNHISAHVNSPEVGVKVNEDLDIVSQSGDLTISFNPTYLIESLKAISKSETVKHFLS 352  
Db 220 SELGELTUTANCTKVGEGKVM-AVNYSGELLEIAFNFFFLDLILKHSKDELVLGIGSD 278

QY 353 VPRPFTLTPGDEESFIQLITPVR 376  
Db 279 SYNPGIIT--DSASGLFVIMPWR 299

RESULT 11  
US-09-134-000C-5040  
; Sequence 5040, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 5040  
; LENGTH: 141  
; TYPE: PRT

; ORGANISM: Enterococcus faecalis  
US-09-134-000C-5040

Query Match 7.1%; Score 131.5; DB 4; Length 141;  
Best Local Similarity 32.6%; Pred. No. 8.5e-05;  
Matches 45; Conservative 16; Mismatches 46; Indels 31; Gaps 5;

QY 71 SPGAILLEASFFINISSLPDISINVKEIQHV-----VLTSG---KSEITLKG 117  
Db 11 SHGSIWTSRPFWGIIRRLPEDMTMEVLNVLGGPFRGGKIPVGGGKREIPPPP 70

QY 118 KDVDQVPRLOEVSTENPLILTKLKSIIAE-----TAPAAASLOESRPILTGHVILSNH 172  
Db 71 PFCGWAPPPRE-----LPGSFGEKNIFTRGGGPMPSRPLTGVHFILENQ 118

QY 173 KDFKAVATDSHRMSORLI 190  
Db 119 K-LLAVATDSHRLSORVI 135

RESULT 12  
US-09-248-796A-15579  
; Sequence 15579, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 15579  
; LENGTH: 977  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-15579

Query Match 6.3%; Score 117.5; DB 4; Length 977;  
Best Local Similarity 20.8%; Pred. No. 0.045;  
Matches 93; Conservative 69; Mismatches 170; Indels 115; Gaps 16;

QY 5 SINRFLFIHALNTRKRAISTKNAIPILSIIKIEV-----TSGVTLTGNGQI---SIE 55  
Db 429 SIFPLTTLTAIAAPTSTAMST-----DLSSSVNIPVETSSNESSITLASSESNAVQTSIE 484

QY 56 NTPVSN-----NAGLLITSPGAILLEASFFINISSLPDISINVKEIE 100  
Db 485 --IPVSSHISANOESSIGESSVNGSTMISSSALESASEFTIAQSSDVESISSSEHSIT 542

QY 101 QHVVLTSG---KSEITLKGKDQVQPRLOEVSTENPLILTKLK-----SIIAETAPA 152  
Db 543 HESSAITSSSTTIETHVTESVVDVSF-----TSDSLFVSETQDPKASSFSIVTPSDSI 596

QY 153 ASLOESRPILTGHVILSNHDKFKAIVATDSHRMSORLITDNTSADLMVLPKSL---R 209  
Db 597 IVNESSIPELSTSVGAVSTSTSSSEISSNSYPFSTHISKVSSSAEPTVTPSESIINI 656

QY 210 EPSAVFTDIE-----TVEFFSPSQILFSEHI 238  
Db 657 SEPAVSTSVQSPSTEQPDMSSVIAESLISSTVSALELSTLMELSHESILAKOSETI 716

QY 239 SFYTRLEGNYPD-----TDRLLMTETFEVVFNTQSLRHAMERAFILS----- 282  
Db 717 N---QSESNIYISETSAVPTAKSTISDSKLTJTWSSSSLVNLESTVSVVMSSEITPIPV 772

QY 283 -----NATQNGTVKLEITQNHISAHVNSPEVGVKVNEDLDIVSQSGDLT---ISFNPTYLIE 336  
Db 773 HTSESATESAESSDIQTFTSTWTTTKSDGFSVETSGVISQSGTSFTTATFPPLY--- 829

```

QY 337 SLKAISKETVKIHLSPVRPFTLTGCD 363
Db 830 -----TSSDITTEFIS---TWTATNSD 848

RESULT 13
US-08-776-265-3
; Sequence 3, Application US/08776265
; Patent No. 6001631
; GENERAL INFORMATION:
; APPLICANT: BLANCHE, Francis
; APPLICANT: CAMERON, Beatrice
; APPLICANT: CROUZET, Joel
; APPLICANT: FAMECHON, Alain
; APPLICANT: FERRERO, Lucia
; TITLE OF INVENTION: No. 6001631el Topoisomerase IV, Corresponding
; TITLE OF INVENTION: Nucleotide Sequences and Uses Thereof
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSER: Dunner, L.L.P.
; STREET: 1300 I. Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/776,265
; FILING DATE: 24-JAN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Einaudi, Carol P.
; REGISTRATION NUMBER: 32,220
; REFERENCE/DOCKET NUMBER: 03806.0394-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4444
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-776-265-3

Query Match 6.2%; Score 115.5; DB 3; Length 800;
Best Local Similarity 18.9%; Pred. No. 0.051;
Matches 89; Conservative 86; Mismatches 174; Indels 123; Gaps 20;

QY 2 IQFSINRFLTHALNTTK-----RAISTKNAIPILSSIKIEVTSGV---TLTG 47
Db 260 IPYEVNKGSLVKRIDELRAADKVDGIVEVREDTDTGLRIALIELKDVNSSEIKNYLYKN 319
QY 48 SNGOISIB-NTIPVSNENAGLITSPGAILLEASFF---INISLDPISINVKIEIQHQ 103
Db 320 SDLOISVNFNVAISDGRPKLM-----GIRQIDSYLNHQIEVAVNRTKFDLDAEKRMHI 375
QY 104 V-----VLTSGKSEITLKGKDQVDQVRLQEVSTENPLILTKLKSIIAE 148
Db 376 VEGLIKALSILDKVIELIRSSKNKDKAKENLIEVYEFTEEQAEAVMLQLYRLNTLTL-- 433
QY 149 TAPRASLOESPIITGVHIVLSNH-----KDFKAVATDSHRMSQRLITLD 193
Db 434 VALEGHEKHELALIKQLRHILDHALNVIKEELNEIKKKFK-----SERLSLIEAIE 488
QY 194 NTSADLMVVLPSK-----SLREPSAVFTDDI-----ET 221

QY 489 ELKIDKEVWVFSEEVILSMTRHGVIKRTSISFNASGVEDICLGDGDSLLKHQEVNTQDT 548
Db 222 VEVFSPSQIILFRSEHISFYTRLLEGNYPD-----TDRLLMTFETFEVVFVNTQSLRHAMER 277
Db 549 VLVFTNKGRLFPVH-----KLARDIRKWLQGHVSQIVPIE-EDEVVINVNEKDFNTD 602
QY 278 AFLISNATONGTVKLEITQNHISAHVNSPEVG-KVNEDLDIVSQ-----SGSD 324
Db 603 AFYVF-ATQNGMIKKSTVPLFKTTTRPNKPLIATKVENDDLSVNRFRKQDLITVITNKG 661
QY 325 LTISFNPTYLIES-LKA--IKSETKVHKHFLSPVRPFTLTGDEBESFIQLIT 373
Db 662 MSLVNTSELSDTGLRAAGVKSINLKVEDF-----VVMTEGVSENDTILMAT 708

RESULT 14
US-09-398-184-3
; Sequence 3, Application US/09398184
; Patent No. 6649394
; GENERAL INFORMATION:
; APPLICANT: BLANCHE, Francis
; APPLICANT: CAMERON, Beatrice
; APPLICANT: CROUZET, Joel
; APPLICANT: FAMECHON, Alain
; APPLICANT: FERRERO, Lucia
; TITLE OF INVENTION: No. 6649394el Topoisomerase IV, Corresponding
; TITLE OF INVENTION: Nucleotide Sequences and Uses Thereof
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSER: Dunner, L.L.P.
; STREET: 1300 I. Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/398,184
; FILING DATE: 17-Sep-1999
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/776,265
; FILING DATE: 24-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Einaudi, Carol P.
; REGISTRATION NUMBER: 32,220
; REFERENCE/DOCKET NUMBER: 03806.0394-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4444
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 800 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
; US-09-398-184-3

Query Match 6.2%; Score 115.5; DB 4; Length 800;
Best Local Similarity 18.9%; Pred. No. 0.051;
Matches 89; Conservative 86; Mismatches 174; Indels 123; Gaps 20;

QY 2 IQFSINRFLTHALNTTK-----RAISTKNAIPILSSIKIEVTSGV---TLTG 47
Db 260 IPYEVNKGSLVKRIDELRAADKVDGIVEVREDTDTGLRIALIELKDVNSSEIKNYLYKN 319

```



RESULT 17  
US-09-248-796A-23131  
; Sequence 23131, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; CURRENT FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 23131  
; LENGTH: 470  
; TYPE: PRT  
; ORGANISM: Candida albicans  
US-09-248-796A-23131

```

Query Match      6.0%; Score 112.5; DB 4; Length 470;
Best Local Similarity 21.3%; Pred. No. 0.041;
Matches 76; Conservative 76; Mismatches 126; Indels 79; Gaps 20;

18 TKRAISTGNAIPILS-SIKIEVTSTGVLTGSGQGISIENTIPVSNENAGLLI---TSPGA 74
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 TKRDPGLSTLSUVINPILSEIKPNPVL-----NQLSFH---PPNPLNTGVLPPPTATAA 53
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

75 ILLEAS-----FFINISSLPDISINVKEIEQHVLTGSKSEITLKGDKVDQVRL 126
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

54 TTTTSSPTNHLHPYFVSLPN-----PQHHPMLDTSQHOETSTDPNQDFYSHL 103
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

127 QEVSTENPLIK---TKLLKSIITAFA-----AASLQSRPILTGVIHIVLNHKDFKAV 178
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

104 -SMPQSTLMLKMYDITTYVRLIETEKYVLLDLANELDWIKKIIPSLCLQISEN----- 156
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

179 ATDHRMSORLITLDNTSADLMVVLPSKSIREFSAFTDDI--EIVEVFPSPQILFRSE 236
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

157 --DSFLDCLMSCSRNTSVNLL-DLTNEQNKWSQKNAPVISERIOOF-----E 203
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

237 HISF-YTRLLEGNPYDTRLTMTTEFTEVVVFNQTS--LEHAMER-----AFLIS 282
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

204 HILISIVLILGLYLNTTKVRLTDYH-KVIFENNQAKLFSHVLKRKTHITITSNKNSAVLT 262
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

293 NATQNGT-VKLEITQNH-ISAHVNSPEVGVKNEDLDIVSQSGDL-----TISFNPTY 333
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

263 NAIQSITMLKPFIDKNYDFEYFNKIQKRGVTDLTLEETIYNSNLSYNPDISYISTF 319
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 18  
US-09-066-046-2  
; Sequence 2, Application US/09066046A  
; Patent No. 6204252  
; GENERAL INFORMATION:  
; APPLICANT: MURPHY, Cheryl  
; STOREY, James  
; BELTZ, Gerald A.  
; COUGHLIN, Richard T.  
TITLE OF INVENTION: CHARACTERIZATION OF GRANULOCYTIC  
; EHRlichia AND METHODS OF USE  
; NUMBER OF SEQUENCES: 65  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HALE AND DORR LLP  
; STREET: 60 State Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: United States  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible

```

1  OPERATING SYSTEM: PC-DOS/MS-DOS
2  SOFTWARE: Patentin Release #1.0, Version #1.30
3  CURRENT APPLICATION DATA:
4  APPLICATION NUMBER: US/09/066,046A
5  FILING DATE: 24-Apr-1998
6  CLASSIFICATION: <Unknown>
7  ATTORNEY/AGENT INFORMATION:
8      NAME: Superko, Colleen
9      REGISTRATION NUMBER: 39,850
10     REFERENCE/DOCKET NUMBER: 106.941.155
11     TELECOMMUNICATION INFORMATION:
12     TELEPHONE: (617) 526-6000
13     TELEFAX: (617) 526-5000
14
15     INFORMATION FOR SEQ ID NO: 2:
16     SEQUENCE CHARACTERISTICS:
17         LENGTH: 619 amino acids
18         TYPE: amino acid
19         STRANDEDNESS: single
20         TOPOLOGY: linear
21
22     MOLECULE TYPE: DNA (genomic)
23     HYPOTHETICAL: NO
24     ANTI-SENSE: NO
25     SEQUENCE DESCRIPTION: SEQ ID NO: 2:
26     US-09-066-046-2

```

```

Query Match      6.0%; Score 112.5; DB 3; Length 619;
Best Local Similarity 20.2%; Pred. No. 0.065;
Matches 81; Conservative 74; Mismatches 158; Indels 87; Gaps 18;

QY      12  IHALNTKRAISKNAIPILSS-----IKLEVTSTGVTLSNGQISIENTIPVSNENAG 66
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Db      215 VNQVNTT--AVQVSESHIIPVSESDTDTVLEDITETITVDGEVGHFSDIASGEHNNDLPA 272

QY      67  LLITSPGAILLEASFPFINIISLPDISI--NVKEIBQHQVVLVTSKGSEITLKGKVDQYP 124
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Db      273 MLLDEADFTMLLANEBESKTLSESPSLEDNVQELG---TLPLQGETVTSSEGNTRRESLP 328

QY      125 RLOEVSTENPLILKTKLLKSLIAETAFAASLOESRRPLTCGVHVLNHNKDFKAVATD--S 182
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Db      329 --TDVSDQSVGVSTDLHAHQSEVETVSEVSTQDS--LSTNI-----SQDSVGVSTDL 378

QY      183 HRMSQRLI----TLDNTSADLMVLPLPSKSLRFSFAVFTDDIETVEVFFSPS----- 229
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Db      379 HSKGVEIVSEGGTQDSLSADFPINTVSESESTDLEA--HSQEVETVSEFTQDSLSTNISQDS 437

QY      230 -----QILFRSEHISPYTRLLEGYDPTRLRLMTFETEVVFTQSLRHAMERAFILSNA 284
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Db      438 VGVSTDLVHRSQVEIVS---EGGTQDS---LSTNISQDSVGVSTDLHAHQSEVETVSEF 491

QY      285 TQMG-----TVKLEI-----TONHISAHVNSPEVGKVNEDL----- 315
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Db      492 TQDSLSTNISQDSVGVSTDLVHRSQVEIVSEGGTQDSLSTNISQDSVG--VSTDLEAHSK 550

QY      316 --DIVSQSGSDLTISPNPYLIESLKAIKSETVKIHFLSP 353
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Db      551 GVEIVSEGGTQDSLSAD-----FPINTVSESESTDLEAHP 585

```

```

1  RESULT 19
2  US-09-248-796A-19154
3  ; Sequence 19154, Application US/09248796A
4  ; Patent No. 6747137
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Keith Weinstock et al
7  ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENC
8  ; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
9  ; FILE REFERENCE: 107196.132
10 ; CURRENT APPLICATION NUMBER: US/09/248,796A
11 ; CURRENT FILING DATE: 1999-02-12
12 ; PRIOR APPLICATION NUMBER: US 60/074,725
13 ; PRIOR FILING DATE: 1998-02-13
14 ; PRIOR APPLICATION NUMBER: US 60/096,409
15 ; PRIOR FILING DATE: 1998-08-13

```

NUMBER OF SEQ ID NOS: 28208  
 SEQ ID NO 19154  
 LENGTH: 1753  
 TYPE: PRT  
 ORGANISM: Candida albicans  
 FEATURE:  
 NAME/KEY: UNSURE  
 LOCATION: (223)  
 OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown  
 US-09-248-796A-19154

Query Match 5.9%; Score 109.5; DB 4; Length 1753;  
 Best Local Similarity 20.8%; Pred. No. 0.69; Indels 121; Gaps 25;  
 Matches 91; Conservative 80; Mismatches 145; Indels 121; Gaps 25;  
 QY 7 NRTFLHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTG-----SNGQI-----SIE 55  
 DB 977 NRTTILE-LSLNPPQINN-LLTDLKADPKQGVFLAVFNNTTNGVWVWSISPIK 1031  
 QY 56 NTPVSNENAGLLITSPGAILLEASFFINIISLP---DISINVKEIE-OHQVLTSGKS 111  
 DB 1032 GTIPM-----LDLTEDGSI-----FQIDNKLPIGNAINVKVKQVDLQHOILVLTARK 1079  
 QY 112 BITLAGKVDQ---VP-RLOEVSTENPLILKTKLKLSIIAETAFASLOESR--PILTVGHVLSNHK 167  
 DB 1080 NFEKEDVRQQTTPARIKIPNHVLV---ELGNVVIASSFTVDALNDYSDKLDHVP- 1135  
 QY 168 VLSNHKDKAVATDSHRMSQRLITLDNTSADLMVVL-----PSKSLREFSAV----- 214  
 DB 1136 ---NVNDY-----VSAKVLDDSTNKRISVSLRTDKAGSKDVINSISDLTRGVQIK 1183  
 QY 215 -PTDIIETVEVFPFSQILF-----RSEHISFYTRLEGNYP-----DTDRLL 256  
 DB 1184 GFVKINSGVYVSLGRSIYALVRVSDLSYLDKQWQNFQKVPVIGKIVNCKQEGRI 1243  
 QY 257 MTEFEDEV-----VENTQSLRHAMERAFILSNATONGT-VKLEITONHISAHVNSPE 307  
 DB 1244 MTKSEVNGELKIMKTFDDQLQGOIFGT--VTSTDFGVFKLDGTVN-VSGLCCHSE 1300  
 QY 308 VGKVED-----LDIVSQSGDLTISFNPTLYL--ESLKAISKSETVKIH 349  
 DB 1301 ISENDVDNVKLFCTGDRVKKVILKIDSEK-RQLSLGKMSAFITDSKEDNDDEVE 1359  
 QY 350 FLSPVRPFTLTGDEEE 366  
 DB 1360 DASEV-----GEEEE 1369

RESULT 20  
 US-09-710-279-546  
 Sequence 546, Application US/09710279  
 Patent No. 6703492  
 GENERAL INFORMATION:  
 APPLICANT: KIMMERLY, WILLIAM JOHN  
 TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
 FILE REFERENCE: PU3480US  
 CURRENT APPLICATION NUMBER: US/09/710,279  
 CURRENT FILING DATE: 2000-11-09  
 PRIOR APPLICATION NUMBER: 60/164,258  
 PRIOR FILING DATE: 1999-11-09  
 NUMBER OF SEQ ID NOS: 4472  
 SOFTWARE: Patent in Ver. 2.1  
 SEQ ID NO 546  
 LENGTH: 532  
 TYPE: PRT  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: Description of Artificial Sequence: synthetic  
 OTHER INFORMATION: amino acid sequence  
 US-09-710-279-546

Query Match 5.9%; Score 109; DB 4; Length 532;  
 Best Local Similarity 19.5%; Pred. No. 0.11;

Matches 76; Conservative 76; Mismatches 160; Indels 78; Gaps 15;  
 QY 19 KRAI-----STKNAIPILSS-----IKIEVTSTGVTLTSGNGO-ISIENTI 58  
 DB 34 KKAIEGIFDIDESKDAINILESIAIDVDEDFLLVKREIFFSSGKSIICRINNQTIVLQDLR 93  
 QY 59 PVSNEAGLLITSPGAILLEASFFINIISLPDISINVKEIEOHOVLTSGHSEITLKGK 118  
 DB 94 KVMQELLDIHGHQETQSLLKQYHLQLLDDYADNOYS-DLLNQYQLSYNQYKN-----RKK 148  
 QY 119 DVDOYPRLOEVSTENPLILKTKLKLSIIAETAFASLOESR--PILTVGHVLSNHK--- 173  
 DB 149 ELEEE-----SAQALLQRLDLMKFQLEET-EASLKEGEVDQLESIDIKRIONSEKLN 202  
 QY 174 ---DFKAVATDSHRMSQRLITLDNTSADLMVVLPSKSLREFSAVFTDIIETVEVFPSPQ 230  
 DB 203 ALNNAHQVLTDESAPDRLYELSNVLOQINDIVPEKFR-----LKEDID----- 247  
 QY 231 ILPRSEHISFYTRLEGNYPDTDRLLMTEPETEVFNQSLRHAMERAFILSNATONGTV 290  
 DB 248 -----QFYMLEDAKHEIYDEKANTEFDEQVLNEYES-----RMNLLNLRK--Y 291  
 QY 291 KLEITONHISAHVNSPEVGVKNEDLDIVSQSGDLTISFNPTLYLIESLKAISKSETVKI-- 348  
 DB 292 GKDITELIAYOSKLANEIDKIEYEQSTQSLREEIKTYNE--VIDICKLSQERRRVAR 349  
 QY 349 ---HFLSPVRPFTLTGDEEESFQLITP 374  
 DB 350 ELRDHIVSEIQNLQMKDANLEISPKPLDEP 379

RESULT 21  
 US-09-248-796A-16697  
 Sequence 16697, Application US/09248796A  
 Patent No. 6747137  
 GENERAL INFORMATION:  
 APPLICANT: Keith Weinstock et al  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA  
 FILE REFERENCE: 107196.132  
 CURRENT APPLICATION NUMBER: US/09/248,796A  
 CURRENT FILING DATE: 1999-02-12  
 PRIOR APPLICATION NUMBER: US 60/074,725  
 PRIOR FILING DATE: 1998-02-13  
 PRIOR APPLICATION NUMBER: US 60/096,409  
 PRIOR FILING DATE: 1998-08-13  
 NUMBER OF SEQ ID NOS: 28208  
 SEQ ID NO 16697  
 LENGTH: 569  
 TYPE: PRT  
 ORGANISM: Candida albicans  
 US-09-248-796A-16697

Query Match 5.8%; Score 107.5; DB 4; Length 569;  
 Best Local Similarity 21.5%; Pred. No. 0.17; Indels 65; Gaps 13;  
 Matches 80; Conservative 56; Mismatches 171; Indels 65; Gaps 13;  
 QY 14 ALNTTKRAISTKNAIPILSSIKIEVTST--GVTLTSGNGQISIENTIPVSNENAGLLITS 71  
 DB 178 ASSSTKSSVSDT--VVNSILLSETSLSDSDTSSTDSISSTTNSDTGNINAGSLHTS 235  
 QY 72 PGAI-----LLEASFFINIISLPDISINVKEIEOHOVLT----- 107  
 DB 236 TASIKELSIQKTGVTLSLSYSLSTKLSSTSDITIELITTELTIEDNEPNTFTSP 295  
 QY 108 SKSEI-----TLGKQVDQYPRLOEVSTENPLILKTKLKLSIIAETAFASLOESRPL 162  
 DB 296 SSHSEIFSSDNLVSKQVDRESTIKTPTDVTVSSLSVHSTEASTATLGNSFSNVA 355  
 QY 163 TGVHVL-----SNHKDFKAVATDSHRMSQRLITLDNTSADLMVVLPSKSLREFSAV 214  
 DB 356 TPLNIATSLRSTSSSNHAT-ESSGTVKSEASAEIIPSPPTSTDNRL-----SYSTEAKGI 411

Qy 215 FTDDIETVEVFPSPQIILFRSHISFYTR-----TDRLLMTFETETEVFN 267  
 Db 412 TVANGSTNNLINESQVAAPTDTSTVLINPVVTFTFDDNSSAAVDQPSKTKSIEESINW 471  
 Qy 268 TQSLRHMERAPFLISNATQNGTVTKLEIQTQNHISAHVNSPEVKGKVNEDLDIVSQGSDLT 326  
 Db 472 PDS-----TNETNNGFI-ATLSQAQVPSSIIHSELITTTAKTTDASMGNDSSA 519  
 Qy 327 ISFNPTYLIESL 338  
 Db 520 SNSQPTLLIOOV 531

RESULT 22  
US-09-134-001C-3172  
; Sequence 3172, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3172  
; LENGTH: 563  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3172

```

Query Match      5.7%, Score 107; DB 3; Length 563;
Best Local Similarity 19.7%; Pred. No. 0.19;
Matches 77; Conservative 74; Mismatches 161; Indels 78; Gaps 15;

Qy 19 KRAT-----SPKNAIPILSS-----IKIEVTSTGVLTGNGQJ-SIENTI 58
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 KKAIEGIFDIDSKDAINILESIAIDVDEDFLLVKREIFSGKSIQRNNQIVITQDLR 124
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 59 PVSNNENAGLLITSPGAILLEASFPFINISSLPDISINVKEIEHQVVLTSCKSEITLKGX 118
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 KWOQLLDIHGHQETQSLLQKVQHLQLLDDVADNQYS-DLLNQVQLSYNQYKN---KRX 179
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 119 DVDQYPLQEVSTENPILIKTKLLKSIIAETAFAASLOERS--PILTVGHIVLSNHK--- 173
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 ELEBLE-----SADQALQLRLDMKQFLESLT-EASLKEGEVQPLESDIKRIQNSKLN 233
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 174 ---DFKAVATSDSHRMSQRLITLNTSADLMVLPKSLRFSVAFVTDIETVEVFFSPSQ 230
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 ALNNAHQVLTDESAPDLRYELSNYLQTDINDIYPEKVR-----LKEDIN----- 278
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 231 ILFRSEHISYTRILLEGYPDTRLMTETETEVFNVTQSIHRAMERAFILISNATQGTV 290
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 279 -----QPYLLLEDAKHEIYDEMANTPEDEQVLNEYES-----RNNLLNNLKRX--Y 322
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 291 KLEITQNHISAHVNSPVGKVNEDLDIVSOGSDLTISFNPTVLTIESLKAIKSETVKI-- 348
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 GKQITELIAYQSKLANEIDKIEYEQSTQSUREIKTYLNE--VIDIGKKLSQERRRVAR 380
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 349 -----HFLSPVRPFTLTGDEEESFIQLITP 374
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 381 ELRDHIYSEIYNLOMKDANLEISFKPLDEP 410
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 23  
US-09-543-681A-6286  
; Sequence 6286, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:

```

; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6286
; LENGTH: 1155
; TYPE: PRT
; ORGANISM: Proteus mirabilis
; US-09-543-681A-6286

```

Query Match	5.7%; Score 106.5; DB 4; Length 1155;
Best Local Similarity	21.1%; Pred. No. 0.68;
Matches	80; Conservative 65; Mismatches 151; Indels 83; Gaps 16;

  

Qy	1	MIOFSINR-----TLFT-----HALNTTKAISTKNAIPILSS 33
Dd	3	VFOYSKLMLMRGSNVFIRLNHFISIFIVILFVLLGNILFTSAKAETVSNNLPTKES 62
Qy	34	IKIEVTSTGVLTCGNSQISIENTIPVSN-ENAGLLITSPCAILLEASFFINISSLPDI 92
Dd	63	IQNQ-----LNLNKRSDLSAEDKLTIGDLEALLLDNIQLEKKADDYNKTEVQLPE- 116
Qy	93	SINVKEIQHQVLTSGKSITLKGKDVOQYPRLQEVSTENPLILKTLLSIIAETAFA 152
Dd	117	-----KRSIQNLQSULKQIA--NKEVDYDROSLE---SLPLATLESUEVEIQLAKA 166
Qy	153	ASLQESRPILTGVHIVLSNHKD--FKAVATDSHRMSQRILTDNSTSADLMVLPSPK---- 206
Dd	167	---QEDLANYSNELIVLQTOPERAQSVLFNNSERLQIRIALNKSSADKAQMRSSVVOLL 223
Qy	207	SLREF-----SAVFDDTIETVEFPSPQILFRSHISFYTRLBGNVPDT 252
Dd	224	QLEGYYLQQONSFKRTLQSNVQLQSLQORDYSSAYIDLQBHAQLLOEILSDKRLDS 283
Qy	253	DRLLMTEFEVTVFNFTOSLRHAMERAFLI SNATONGTV--KLETTONHSIHAVNUSPEVGK 310
Dd	284	SEEVAKAQTAELSN-QAIKN---NAPYLQAADINKKLSDKLIITTO-----NNNELNR 333
Qy	311	-----VNBDLDIVSQSGDL 325
Dd	334	HSLMVKNRLDRAICQSERNL 352

```

RESULT 24
US-09-700-227-2
; Sequence 2, Application US/09700227
; Patent No. 6706494
; GENERAL INFORMATION:
; APPLICANT: Ruelle, Jean-Louis
; TITLE OF INVENTION: BAS8028 Polypeptides and Polynucleotides
; TITLE OF INVENTION: Encoding Therefor From Moraxella Catarrhalis
; FILE REFERENCE: BM45319
; CURRENT APPLICATION NUMBER: US/09/700,227
; CURRENT FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: PCT/EP99/03263
; PRIOR FILING DATE: 1999-05-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1726
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-700-227-2

```

Query Match 5.6%; Score 104.5; DB 4; Length 1726;  
Best Local Similarity 18.2%; Pred. No. 2;  
Matches 72; Conservative 61; Mismatches 147; Indels 115; Gaps 13;



QY 6 INRTLFHANTTKRAISTKNAIPILSSIKIEVTSTG-----VTLTSGNGQISIENT 57  
DB 581 IARTTKSPILNSKNAQVLNGLSLQTDRLALTNSTNHNLAITGDISLVQOQKLTQNT 640  
QY 58 IPVSNENAGLLITSPGAILLEASFFINISSLPDISINVKETEHOHVLTSGSKSITLKG 117  
DB 641 NPVTNGN-----QAHYLN-----AQGNLTGQSEVQLOG 671  
QY 118 KDQVQPRQEVSTENPLIL-----KTKLKLSIIAETAFASIQESRPILTGVHIVLS 170  
DB 672 VNLKSKNNLTAVATQGVMDGVKREFSNKQSGDQKALEKLEA-----ARLSIT 725  
QY 171 NHHKFKAVATDSHRMSORLITLNTSADLMVVLPSKSLREFSAVFTDDIETVEVFFSPQ 230  
DB 726 TRAD--AFTNTKRIENQILNLSNSTGFI-----HTPTT 758  
QY 231 ILFRSEHISFYTR-----LLEGN-----YPTDRLMLTFEVEVFN 267  
DB 759 LVSQAGDIGYAKQILSGALLNAVSGNITLEAQGVLPFAQOQTKSQSGSSNQSDAE 818  
QY 268 TQSLRHAMERAFILSNATQGVTKLEITQ-----NHSIAHVNSPEVGVKNEDLDIVSQSGS 323  
DB 819 TQT-----DKAASIIITGSON-----RYEIGQDGDHRLLSFGSPSMLTAGRDIDIRAKAA 871  
QY 324 DLTISFNPTLYLESIAKSETVKI-----HFLSPV 354  
DB 872 TDQDETPTLVLOGTAKANGAIDIRANHHLLDV 906

## RESULT 25

US-09-107-532A-5248  
; Sequence 5248, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD/ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 5248:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 564 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:

; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...564  
; SEQUENCE DESCRIPTION: SEQ ID NO: 5248:  
US-09-107-532A-5248  
Query Match 5.6%; Score 103.5; DB 4; Length 564;  
Best Local Similarity 22.2%; Pred. No. 0.41;  
Matches 82; Conservative 46; Mismatches 119; Indels 123; Gaps 19;  
QY 27 AIPILSSIKIEVTSTGVTLTGSGNGQISIENTIPVSNENA-----GLLI 69  
DB 82 ALPYLLS-KISVPVFGTETLTIELAKLNSR-----NEAKGFKDHFVVDHEITDFGETV 135  
QY 70 TSPGAILLEASFFINISSLPD-ISINVKETIEHQVLTSGKSEITLKGKDVQDQYPRLOE 128  
DB 136 VS-----PFTTHITPDSVGINLKTREGSVYVTDGDFKQDTAIPMYQTDYARLAE 185  
QY 129 VSTENPLILKTK-----LKSIIAETAF-----AASLOESRPILT 163  
DB 186 ICKEGLVALLSESANAENPQPVASQAIADEVFTDIKYWEGR11VACVASNLQRVQVLD 245  
QY 164 GVH-----IVLSNKHDKFKAATDSHRMSORLITLNTSADLMVVLPSKSLREFSAVFTDD 218  
DB 246 AAQKAKRVLTG-ODFERII-----RTAMRLEKQLPSEDLVLTL--KDMKN----- 290  
QY 219 IETVEVFFSPQILF-----RSEHISFYTRLEGNYPTDRLMLTFEVEVFNQSLRH 273  
DB 291 -----YDPELIILETGRMGEPIKSLQKWANG-----THRNKTE-EGDLVYITTPSI 338  
QY 274 AMERAF-----LISNATQGVTKLEITONHISAHVNSPEVGVKNEDLDIVSQSGSLTIS 328  
DB 339 AMETIVAKTEDIIVRA--GGTVKLISENMRVSGHANP-----NDLOLM 379  
QY 329 FN---PTVLI 335  
DB 380 INLMKPTTYFV 389  
RESULT 26  
US-09-463-402-6  
; Sequence 6, Application US/09463402  
; Patent No. 6596510  
; GENERAL INFORMATION:  
; APPLICANT: Lubitz, Werner  
; APPLICANT: Resch, Stephanie  
; TITLE OF INVENTION: Secretion of Carrier-bound Proteins into the Periplasm  
; FILE REFERENCE: 05649059  
; CURRENT APPLICATION NUMBER: US/09/463,402  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: DE19732829.6  
; PRIOR FILING DATE: 1997-07-30  
; PRIOR APPLICATION NUMBER: PCT/EP98/04723  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 920  
; TYPE: PRT  
; ORGANISM: Bacillus stearothermophilus  
US-09-463-402-6  
Query Match 5.6%; Score 103.5; DB 4; Length 920;  
Best Local Similarity 20.8%; Pred. No. 0.91;  
Matches 77; Conservative 59; Mismatches 146; Indels 89; Gaps 16;  
QY 16 NTTKRAISTKNAIPILSSIKIEVTSTGVTLT-TGSGNGQISIENTIPVSNENAGLLITSPGA 74  
DB 447 DATVRSLN-----PIIATAINGSELLVTANAGQSGKASFEVTPKDNKTRTFTVDVKKDP 501  
QY 75 ILLEASFFINISSLPDISINVKETIEHQVLTSGKSEITLKGKDVQDQYPRLOEVSTENP 134

Db 502 VLQDIKVDATSV-KLSDEAVGGEVE-----GWNQKTIKVSADVQYGEIKGTGK 552  
Qy 135 LILKTK-----LLKSIATAFAASLQESRPILTGVHIVLSNH-----KDFKAVAT 180  
Db 553 VVTVTNTTEGLVKNVNSDNTIDF---DSGNSATDQFVVVATKDKIVNGKVEVKFKN-AS 608  
Qy 181 DSHRMSQRLITL-----DNTSADLMVVLPSKSLRFSAVFTDDIETVEVFFSPSILF 233  
Db 609 DTTPTSTKTIIVNVVVKADATPVGLDIVARS-----EIDVNPATSTADVDF-----I 658  
Qy 234 RSEHISFYRLLGN-----YPTDRLLMTFETEVFNTQSLRHAMERAFILSNATQNG 288  
Db 659 NFESVEIYTLDSNGNRLKVKVTPATTGLTNDYVEVNGVNLQFKGNDELTL-----TSSS 714  
Qy 289 TVKLEITQNHISAHV-----NSPEVGKVNEDLDIVSQSGDLTIS----- 328  
Db 715 TVNVDVTADGITKRIPIVKYINSASVPASATVATSPVTVKLN-----SSNDLTFEELI 767  
Qy 329 ---FNPTVLIIE 336  
Db 768 FGVIDPTQLVK 778

## RESULT 27

US-09-889-572-4  
; Sequence 4, Application US/09889572  
; Patent No. 6610517  
; GENERAL INFORMATION:

; APPLICANT: Lubitz, Werner  
; TITLE OF INVENTION: COMPARTMENTALIZATION OF RECOMBINANT POLYPEPTIDES IN HOST CELLS  
; FILE REFERENCE: 100564-00070  
; CURRENT APPLICATION NUMBER: US/09/889,572  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: PCT/EP00/00686  
; PRIOR FILING DATE: 2000-01-28  
; PRIOR APPLICATION NUMBER: DE 199 03 345.5  
; PRIOR FILING DATE: 1999-01-28  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 921  
; TYPE: PRT  
; ORGANISM: Bacillus stearothermophilus  
US-09-889-572-4

Query Match 5.6%; Score 103.5; DB 4; Length 921;  
Best Local Similarity 20.8%; Pred. No. 0.91;  
Matches 77; Conservative 59; Mismatches 146; Indels 89; Gaps 16;

Qy 16 NTKRAISTKNAIPILSSIKIEVTSTGYTL-TGSGNGQISIENTIPVSNENAGLLITSPGA 74  
Db 448 DATVRSIN-----PIATAANGSELLVATNAGQGGKASFVTFKDNTRFTTVDVKKDP 502  
Qy 75 ILLASFFINISSLIPDISINVKIEQHVLTGSKSEITLKGKDVQDQYPRLOEVSTENP 134  
Db 503 VLQDIKVDATSV-KLSDEAVGGEVE-----GWNQKTIKVSADVQYGEIKGTGK 553  
Qy 135 LILKTK-----LLKSIATAFAASLQESRPILTGVHIVLSNH-----KDFKAVAT 180  
Db 554 VVTVTNTTEGLVKNVNSDNTIDF---DSGNSATDQFVVVATKDKIVNGKVEVKFKN-AS 609  
Qy 181 DSHRMSQRLITL-----DNTSADLMVVLPSKSLRFSAVFTDDIETVEVFFSPSILF 233  
Db 610 DTTPTSTKTIIVNVVVKADATPVGLDIVARS-----EIDVNPATSTADVDF-----I 659  
Qy 234 RSEHISFYRLLGN-----YPTDRLLMTFETEVFNTQSLRHAMERAFILSNATQNG 288  
Db 660 NFESVEIYTLDSNGNRLKVKVTPATTGLTNDYVEVNGVNLQFKGNDELTL-----TSSS 715  
Qy 289 TVKLEITQNHISAHV-----NSPEVGKVNEDLDIVSQSGDLTIS----- 328  
Db 716 TVNVDVTADGITKRIPIVKYINSASVPASATVATSPVTVKLN-----SSNDLTFEELI 768

Qy 329 ---FNPTVLIIE 336  
Db 769 FGVIDPTQLVK 779

## RESULT 28

US-09-134-000C-4971  
; Sequence 4971, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/055,778  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4971  
; LENGTH: 1233  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-4971

Query Match 5.5%; Score 103; DB 4; Length 1233;  
Best Local Similarity 20.5%; Pred. No. 1.6;  
Matches 86; Conservative 75; Mismatches 115; Indels 144; Gaps 24;

Qy 46 TGSNGQI--STENTIPVSNENAGLLIT-SPGAIL-----LEASFFINISSLIPDISINVK 97  
Db 263 TGLGGEIIFYNLDRTL-TGNQTLLELTETTPGAVFGKQDNLEPQVFSY-----DVDINGQ 315  
Qy 98 EIEQHVLTGSKSEITLKGKDVQDQYPRLOEVSTENPILKTLKLSIIAETAFAASLQE 157  
Db 316 ILPETQTLTTPGK-DYTL-----SDNSLGRITAVTPVNNQKQVLSINR 359  
Qy 158 SRPILTGHVILSNHDKFKAATDSHRMSQRLITLNTSADLMVVLPSKSLRFSAVT 217  
Db 360 T-----IYLESADY-----NYLYSQOYPTTKIGSISLKSSTGTGKOTDTFAKTSQ 405  
Qy 218 DIETV--EVFPSPQIILFRSE--HISFYTRLLGNYPDTDRLEMTETETEVWF---NTQ 269  
Db 406 TSKVIADREMSYISFQSKGYVIYGLTE-----TKVGOQIVLESTNGQ 454  
Qy 270 SLRHAMERAF-----LISNATQNGTVKLEIT-----QNHISA 301  
Db 455 EIKNPKFTAYGLYENVKLEDFDIKTEGGKLTLTATKDSYLINISDLTMDFKDKDINL 514  
Qy 302 HVNSPEVKVNEDLDIVS-----QSG---SDLTISP----- 329  
Db 515 SLSTFVIGP-NKAIQLVSDQYIEPISVNVPLNAETAAGNYDNGAYSRTTTSVNGSKEK 573  
Qy 330 -----NPTYLIESLKAIKSETVKIHFSPV-RPFTLTGDEEESFQILITPVRTN 378  
Db 574 PIONLEIKVHPNVL--SLRATK-----EIVFYKLGTDYVTP-TSDGSVIKFTTPI-TN 625

## RESULT 29

US-09-071-035-404  
; Sequence 404, Application US/09071035  
; Patent No. 6448043  
; GENERAL INFORMATION:

; APPLICANT: Gil H. Choi  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 496  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland

COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/071,035  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: A. Anders Brookes  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB369P2  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 404:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1416 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-071-035-404

Query Match 5.5%; Score 103; DB 4; Length 1416;

Best Local Similarity 20.5%; Pred. No. 2.1;  
Matches 86; Conservative 75; Mismatches 115; Indels 144; Gaps 24;

QY 46 TGSNGQI--SIENTIPVSNENAGLLIT--SPGAIL-----LEASFFINISSLPDISINVK 97  
DB 232 TGLDGEIFYNLDRTL--TGNQTLLELTTPGAVFGKQDNLEPQVFSY-----DVDINGQ 284  
QY 98 EIEQHVLTGSKSEITLKGKDVQYPRLQEVSTENPLILTKLKLSIIAETAFASLOE 157  
DB 285 ILPTOTLLTPCK--DYTL-----SDNSLGRIAVTVPNMNOQKAYSLSINR 328  
QY 158 SRPILTGVIHVLNKHDFKAVATDSHRMSQRLITLDNTSADLMVVLPSKSLREFSAVTD 217  
DB 329 T-----IYLESASDY-----NYLSYQQYPTTKIGSISLKSTGKTQTTDFTAKTSQ 374  
QY 218 DIETV--EVFPSPQILFRSE---HISFYTRLLEGNYPDTRLMLTETETEVVF---NTQ 269  
DB 375 TSKVIADREMRMSYISFQSKGYVYTYGLTE-----TKVQQQIVLESTNGQ 423  
QY 270 SLRHAMERAF-----LISNATQNGTVKLEIT-----QNHISA 301  
DB 424 EIKPKFTAYGLYENVKLEDYFDIKTEGGKLTLTATKDSYLRLINISDLTMDFKKDINL 483  
QY 302 HVNSPEVGKWNEDDIVS-----QSG---SDLTISF----- 329  
DB 484 SLSTEVIGP--NKAIQLVSDQYIEPISVVNPLNAETAWGNYDQNGAYSRRITVSVNGSKEK 542  
QY 330 -----NPTYLIESLKAISKETVKIHLFSPV--RPFTLTPGDEESFQILITPVRTN 378  
DB 543 PQNLEIKVGHENYL--SLRATK-----EIFYFKLGTDTYVTP--TSDGSVIKFTTPI--TN 594

RESULT 30

US-09-071-035-402  
Sequence 402, Application US/09071035  
Patent No. 6448043  
GENERAL INFORMATION:  
APPLICANT: Gil H. Choi  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 496  
CORRESPONDENCE ADDRESS:  
ADDRESSES: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/071,035  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: A. Anders Brookes  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB369P2  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 402:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1448 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-071-035-402

Query Match 5.5%; Score 103; DB 4; Length 1448;

Best Local Similarity 20.5%; Pred. No. 2.1;  
Matches 86; Conservative 75; Mismatches 115; Indels 144; Gaps 24;

QY 46 TGSNGQI--SIENTIPVSNENAGLLIT--SPGAIL-----LEASFFINISSLPDISINVK 97  
DB 259 TGLDGEIFYNLDRTL--TGNQTLLELTTPGAVFGKQDNLEPQVFSY-----DVDINGQ 311  
QY 98 EIEQHVLTGSKSEITLKGKDVQYPRLQEVSTENPLILTKLKLSIIAETAFASLOE 157  
DB 312 ILPTOTLLTPCK--DYTL-----SDNSLGRIAVTVPNMNOQKAYSLSINR 355  
QY 158 SRPILTGVIHVLNKHDFKAVATDSHRMSQRLITLDNTSADLMVVLPSKSLREFSAVTD 217  
DB 356 T-----IYLESASDY-----NYLSYQQYPTTKIGSISLKSTGKTQTTDFTAKTSQ 401  
QY 218 DIETV--EVFPSPQILFRSE---HISFYTRLLEGNYPDTRLMLTETETEVVF---NTQ 269  
DB 402 TSKVIADREMRMSYISFQSKGYVYTYGLTE-----TKVQQQIVLESTNGQ 450  
QY 270 SLRHAMERAF-----LISNATQNGTVKLEIT-----QNHISA 301  
DB 451 EIKPKFTAYGLYENVKLEDYFDIKTEGGKLTLTATKDSYLRLINISDLTMDFKKDINL 510  
QY 302 HVNSPEVGKWNEDDIVS-----QSG---SDLTISF----- 329  
DB 511 SLSTEVIGP--NKAIQLVSDQYIEPISVVNPLNAETAWGNYDQNGAYSRRITVSVNGSKEK 569  
QY 330 -----NPTYLIESLKAISKETVKIHLFSPV--RPFTLTPGDEESFQILITPVRTN 378  
DB 570 PQNLEIKVGHENYL--SLRATK-----EIFYFKLGTDTYVTP--TSDGSVIKFTTPI--TN 621

RESULT 31

US-09-134-001C-5119  
Sequence 5119, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

;; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

;; FILE REFERENCE: GTC-007

;; CURRENT APPLICATION NUMBER: US/09/134,001C

;; CURRENT FILING DATE: 1998-08-13

;; PRIOR APPLICATION NUMBER: US 60/064,964

;; PRIOR FILING DATE: 1997-11-08

;; PRIOR APPLICATION NUMBER: US 60/055,779

;; PRIOR FILING DATE: 1997-08-14

;; NUMBER OF SEQ ID NOS: 5674

;; SEQ ID NO 5119

;; LENGTH: 395

;; TYPE: PRT

;; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-5119

Query Match 5.5%; Score 102.5; DB 3; Length 395;

Best Local Similarity 20.3%; Pred. No. 0.28;

Matches 79; Conservative 59; Mismatches 117; Gaps 17;

QY 30 ILSSIKIEVTSTGVLTSNGQISIENTIPV-SNENAGLITSPGAILLEASFFINI1SS 88

Db 28 ILSAQLPMVGHRSSTDFEIAEAPKGLXPFGSKNEVLITSSGTSVLEASM----- 80

QY 89 LPDISINVKIEBQHVLTSGKSEITLK--GRVDQVPRLOEVSTENPLILK---TKLLK 143

Db 81 -----LNIANPDHIVIIVSGAFGNRFKQIAQTYNHVHVVDNNGEAVIVDDFTYLLKQ 135

QY 144 SIIAETAFASLOESRPILTGV-HIVLSNHKDFKAVATDSHRMSORLITLNT----- 195

Db 136 LNPVTAFTQCETS---TGVHVPV-----HQLGHAKAFDNSLYFIYVDG 178

QY 196 -----SADLMVLPKSLREFS-----AVPTD-----DIETVEVF----- 225

Db 179 VSCIGAVDVLTKIDVLVSGQKAIMLPGLAFVAVSDRAKKFADVKTPRFYDLNKK 238

QY 226 -----FSPQILFRSHISFYTRLLEGNYPTDRLMLTEFETEVVNTQSLRHA 274

Db 239 YIKSQEQNSTPTPNVGLFRG--INAYVELVK-----KEGLNHV 275

QY 275 MERAFISNATQGTVKLEI-----TONHISAHVNSPEVGKVEDDIVSQSGDLTISF 329

Db 276 ISRHFKIRNALRAUKALELELLVKDDAHASPTVTS-FVPRNQBELNLIK---NOLKSQF 331

QY 330 NPTYLIESLKAIKSETVKHFLSPVRPFTL 359

Db 332 NIT-IAGQGHKGGQILRIGHMGKISPFDI 360

RESULT 32

US-09-248-796A-23039

;; Sequence 23039, Application US/09248796A

;; Patent No. 6747137

;; GENERAL INFORMATION:

;; APPLICANT: Keith Weinstock et al

;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

;; FILE REFERENCE: 107196.132

;; CURRENT APPLICATION NUMBER: US/09/248,796A

;; CURRENT FILING DATE: 1999-02-12

;; PRIOR APPLICATION NUMBER: US 60/074,725

;; PRIOR FILING DATE: 1998-02-13

;; PRIOR APPLICATION NUMBER: US 60/096,409

;; PRIOR FILING DATE: 1998-08-13

;; NUMBER OF SEQ ID NOS: 28208

;; SEQ ID NO 23039

;; LENGTH: 461

;; TYPE: PRT

;; ORGANISM: Candida albicans

US-09-248-796A-23039

Query Match

Best Local Similarity 18.6%; Pred. No. 0.36;

Matches 64; Conservative 76; Mismatches 154; Indels 51; Gaps 11;

QY 19 KRAISTKNAIPILSIIKIEVTSTGVLTSNGQISIENTIPVSNENAGLITSPGAIL-L 76

Db 5 KSVLPVKALPVGKALPSEVSSSS-EVSSSQVISSEVSSSSSE---VVSSSEVSSSS 59

QY 77 LEASFFINISSLPDIS-----INVKEIEHQHVLTSGKSEITLKGD 119

Db 60 SEVSSSEVSSSSSEVSSSSSEVSSSQVTSSEIVSSSEVSSSEVSSSEVSSSSSE 119

QY 120 VDOYPRL---QEVSTENPLIKTKLLKSLIAETAFASLOESRPILTGVHIVLSNHK-- 173

Db 120 VSSSEVSSSSSEVSSSSSEVSSSSSEVSSSSSEVSSSQVTSSEIVSSSEVSSSSSEVSS 177

QY 174 -DFKAVATDSH-RMSQRLITLNTSADLMVLPKSLREFAVFTDDITV----- 222

Db 178 SSSEVSSSSSEVSSSSSEVSSSQVTSSEIVSSSEVSSSSSEVSSSSSEVSSSS 237

QY 223 -EVFPSPQILFRSE---HISFVTRLLEGNYPTDRLMLTEFETEVVNTQSLRHAWE 276

Db 238 SEVSSSEVSSSSSEVSSSSSEVSSSQVTSSEIVSSSEVSSSSSEVSSSSSEVSSSSSE 297

QY 277 RAFLISNATQGTVKLEITONHISAHVNSPEVGKVEDDIVSQS 321

Db 298 VSSSSSEVSSSSSEVSSSSSEVSSSQVTSSEIVSSSEVSSSSSEVSSSSSEVSSSSSE 340

RESULT 33

US-08-480-604A-10

;; Sequence 10, Application US/08480604A

;; Patent No. 5736139

;; GENERAL INFORMATION:

;; APPLICANT: KINK, JOHN A.

;; APPLICANT: THALLEY, BRUCE S.

;; APPLICANT: FADHYE, NISHA V.

;; APPLICANT: FIRCA, JOSEPH R.

;; APPLICANT: STAFFORD, DOUGLAS C.

;; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND

;; PREVENTION OF C. DIFFICILE DISEASE

;; NUMBER OF SEQUENCES: 32

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: MEDLEN & CARROLL, LLP

;; STREET: 220 MONTGOMERY STREET, SUITE 2200

;; CITY: SAN FRANCISCO

;; STATE: CALIFORNIA

;; COUNTRY: UNITED STATES OF AMERICA

;; ZIP: 94104

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/480,604A

;; FILING DATE: 07-JUN-1995

;; CLASSIFICATION: 424

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 08/422,711

;; FILING DATE: 14-APR-1995

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 08/405,496

;; FILING DATE: 16-MAR-1995

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 08/329,154

;; FILING DATE: 25-OCT-1994

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 08/161,907

;; FILING DATE: 02-DEC-1993

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/985,321

;; FILING DATE: 04-DEC-1992

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/429,791

;; FILING DATE: 31-OCT-1989



Patent No. 6365158  
GENERAL INFORMATION:  
APPLICANT: Williams, James A.  
APPLICANT: Kink, John A.  
TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES  
TITLE OF INVENTION: OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE  
TITLE OF INVENTION: DISEASE  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/957,310  
FILING DATE: 23-OCT-1997  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/329,154  
FILING DATE: 24-OCT-1994  
APPLICATION NUMBER: US 08/161,907  
FILING DATE: 02-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/985,321  
FILING DATE: 04-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/429,791  
FILING DATE: 31-OCT-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Ingolia, Diane E.  
REGISTRATION NUMBER: 40,027  
REFERENCE/DOCKET NUMBER: OPDH-01121  
TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-957-310-10

Query Match          5.5%; Score 102; DB 3; Length 2366;
Best Local Similarity 20.9%; Pred. No. 6;
Matches 72; Conservative 61; Mismatches 139; Indels 72; Gaps 17;

QY      30 ILSSIKI-----EVTSTGTVLTGNGOISIENTIPVSNAGLLITSPCAILLEA 79
       ||| : | | : | | : | | : | | : | | : | | : | | : | | : |
Db      1371 ILTSLGIENKILNLSHEINFPSG-EVNGNGFVSLTFSL-LEGINALIEVD----LLSK 1423

QY      80 SPFINIISSLPDISINVKEIQHVVLTSQKSITILKGKDVOQYPRQLQEVSTENPLI--- 136
       || : | : | : | : | : | : | : | : | : | : | : | : |
Db      1424 SYKLILSGELKILMLNSNIQQ-KIDYIGNSELQ--KNII-PYSFVDSSEKNGFINGS 1478

QY      137 LXTKLLKSIIAETAFAAS--LOESRPILTGCVHIVLNHKDFKAVATDSHRMSORLITLDN 194
       || : | : | : | : | : | : | : | : | : | : | : | : |
Db      1479 TXEGLFVSLEPDLVVILSKVYMDDSKP---SFGYYSNNLKDVKITDKNVNIIITYYLKOD 1535

QY      195 TSADLMVVLPSKSLRFSAVFDDDIETVEVFVFPSPQILFRSEHSIFVTRLEGNYPDTDR 254
       || : | : | : | : | : | : | : | : | : | : | : | : |
Db      1536 IKISLSLTLDQEKTIKLVSHLDESGVAII-----LKFWNR--KGNWTNTSDS 1580

QY      255 LL-----MTEFEDEVFNFTQSRLHAMERAPFLISNATQNGT-----VKLEI 294
       || : | : | : | : | : | : | : | : | : | : | : | : |
Db      1581 LMSFIIESNMKI GFVN FLOSNIKFIL DANFII SGTT SI GOFEPICDENNDNIOPVFI KFT 1640

```

US-08-957-310-10  
; Sequence 10, Application US/08957310

QY 295 TQNHSAHVNSPE--VGKVNEDLDIVSQSG--SDLTISFNPTYL 334  
Db 1641 LETNYTLYVGNQNMVPEYDLD---DSGDISSTVINFQSKYL 1681

RESULT 37  
US-10-011-366-10  
; Sequence 10, Application US/10011366  
; Patent No. 6573003  
; GENERAL INFORMATION:  
; APPLICANT: Williams, James A.  
; TITLE OF INVENTION: IDENTIFICATION OF NEUTRALIZING EPITOPES  
; OF TOXIN A AND TOXIN B FOR THE TREATMENT OF C. DIFFICILE  
; DISEASE  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/10/011,366  
; FILING DATE: 16-NOV-1993  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/957,310  
; FILING DATE: 23-OCT-1997  
; APPLICATION NUMBER: US/08/329,154  
; FILING DATE: 24-OCT-1994  
; APPLICATION NUMBER: US/08/161,907  
; FILING DATE: 02-DEC-1993  
; APPLICATION NUMBER: US/07/985,321  
; FILING DATE: 04-DEC-1992  
; APPLICATION NUMBER: US/07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ingolia, Diane E.  
; REGISTRATION NUMBER: 40,027  
; REFERENCE/DOCKET NUMBER: OPHD-01121  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2366 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-10-011-366-10

Query Match 5.5%; Score 102; DB 4; Length 2366;  
Best Local Similarity 20.9%; Pred. No. 6;  
Matches 72; Conservative 61; Mismatches 139; Indels 72; Gaps 17;

QY 30 ILSIXI-----EVTSTGVTLTGNGQISIENTIPVSNENAGLITSPGAILLEA 79  
Db 1371 ILSLSEENKIILNSHEINFSG-EVNGSNGFVSLTFSI-LEGINAIIEVD-----LLSK 1423  
QY 80 SFPIINISLPDISINVKEHQHVLTSGSKSITLTKGDKVDQVYPLQVSTENPLI--- 136  
Db 1424 SYKLLISGELKILMUNSHIQQ-KIDYIGFNSBLQ---KNI-PYSVDFSEKENGFGINGS 1478  
QY 137 LKTKLLKSIATAFAAAS--LOESRPILTGVHVLNSHKDFKAVATDSHRMSQRLITLON 194

Db 1479 TKEGLFVSELPDVVLISKVYMDSKP---SFGYYNNLKDVKVITKDNVNIITGYLKKDD 1535  
QY 195 TSADLMVVLPSKSLREFSAVFTDDTETVEVFSPSQILFRSEHISFYTLLEGNYPDTR 254  
Db 1536 IKISLSLTLODEKTIKLSVHLDESQVAEI-----LKPMNR--KGNTNTSDS 1580  
QY 255 LL-----MTEFETEVVNTQSLRHAMERAFILSNATQNGT-----VKLEI 294  
Db 1581 LMSFLESMNIKISIFVNFQSNIKFILDANFIISGTTISGQFEFICDENDNIQPYFIKNT 1640  
QY 295 TQNHSAHVNSPE--VGKVNEDLDIVSQSG--SDLTISFNPTYL 334  
Db 1641 LETNYTLYVGNQNMVPEYDLD---DSGDISSTVINFQSKYL 1681

RESULT 38  
US-09-084-517-10  
; Sequence 10, Application US/09084517  
; Patent No. 6613329  
; GENERAL INFORMATION:  
; APPLICANT: Kink, John A.  
; APPLICANT: WILLIAMS, JAMES A.  
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND  
; PREVENTION OF C. DIFFICILE DISEASE  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL  
; STREET: 220 MONTGOMERY STREET, SUITE 2200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: UNITED STATES OF AMERICA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/084,517  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/  
; FILING DATE: 16-MAR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/329,154  
; FILING DATE: 25-OCT-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/161,907  
; FILING DATE: 02-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/985,321  
; FILING DATE: 04-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/429,791  
; FILING DATE: 31-OCT-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CARROLL, PETER G.  
; REGISTRATION NUMBER: 32,837  
; REFERENCE/DOCKET NUMBER: OPHD-01610  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2366 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-084-517-10

Query Match 5.5%; Score 102; DB 4; Length 2366;



Best Local Similarity 20.9%; Pred. No. 6; Matches 72; Conservative 61; Mismatches 139; Indels 72; Gaps 17;  
QY 30 ILSSIKI-----EVTSGVLTGSGNGOISIENTIPVSNENAGLLITSPGAILLEA 79  
Db 1371 ILUSTIENKILMSNHIQ--KIDYIGFNSLQ---KNI-PYSFVDSGKNGFINGS 1478  
QY 80 SFFINIISLPDISINVKEIQHQVLTSGKSEITLKGKDVDPQYPRLOEVSTENPLI--- 136  
Db 1424 SYKLLISGELKILMSNHIQ--KIDYIGFNSLQ---KNI-PYSFVDSGKNGFINGS 1478  
QY 137 LTKLLKSIITATAFAS--LOESRILTGVHIVLSNHKDFKAVATDSHRMSQRLITDN 194  
Db 1479 TKEGLFSELPLDVLISKYMDSKP---SFGYYSNNLKDVKVITKDNVILTYGLKDD 1535  
QY 195 TSADLMVPLPSKSLREFSAVETDITVEVFPSPQILFRSEHSFYTRLLEGNYPDTR 254  
Db 1536 IKISLSLTQDEKTKILMSVHLDGSAEI-----LKPMMR--KGNTNTSDS 1580  
QY 255 LL-----MTEFETEVFNTQSLRHAMERAFILISNATQGT-----VKLEI 294  
Db 1581 LMSPLSMNIIKIFVFLQSNKIFLDANFIISGTTISGQFEFICDENDNIQPYFIKNT 1640  
QY 295 TQNHISAHVNSPE--VGKNEBIDIVSQSG--SDLTISFNPTYL 334  
Db 1641 LETNYTLVGNRQNMIVEPNYDLD--DSGDISSTVINFSQKYL 1681

RESULT 39  
US-09-538-092-326  
; Sequence 326, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CuraPatSeqFormatter Version 0.9  
; SEQ ID NO 326  
; LENGTH: 960  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (0)-(0)  
; OTHER INFORMATION: Polypeptide Accession Number YGL238W  
US-09-538-092-326

Query Match 5.4%; Score 101.5; DB 4; Length 960;  
Best Local Similarity 24.4%; Pred. No. 1.5; Matches 65; Conservative 44; Mismatches 94; Indels 63; Gaps 14;  
QY 101 QHQVLTSGKSI-----TLKGKDVDPQYPRLOEVSTENPLIKYTKLKSIIATAFAA 153  
Db 254 EHASVLIKVKSSIQELVQLYTRYEDVFG-PMINEF-----IQITWNLLTSSNQPKYDI 307  
QY 154 SIQESRPLTGVHIVLSNHKDFKAVATDSHRMSQRLITDNLSADLMVPLPSKSLREFSA 213  
Db 308 LVSKLSLFLATVRI---PKYFEIENNES-----AMNNITE---QIILPNVTLRE--- 351  
QY 214 VETDDIETVEVFPSPQILFRSEHSFYTRLLEGNYPDTRLLMTEF-----ETEWFVN 267  
Db 352 ---EDVE-----LFEDDPIEYIRRDLEGGSDTTRREACTDFLKEKKEVNLVT 397  
QY 268 TQSLRHAMERAFI---ISNATQGTGK-----LEITQNHISAHVNSPEVGKVNEDLDIV-- 318

Db 398 NIFLAHM--KGFVDQYMSDPSPKQKFKDLYLYLFTALAINGNITNAGVSSNTNLLNVVDF 455  
QY 319 --SQSGSLTISFNFETYLIESLKAIAK 342  
Db 456 FTKEIAPDLT-SNNIPHILRVDAIK 480  
RESULT 40  
US-09-543-681A-4674  
; Sequence 4674, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETTON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 4674  
; LENGTH: 385  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-4674

Query Match 5.4%; Score 101; DB 4; Length 385;  
Best Local Similarity 20.3%; Pred. No. 0.38; Matches 74; Conservative 58; Mismatches 123; Indels 110; Gaps 17;  
QY 1 MIQFSINRSL-----FIHALNTTKRAISTKNAIPILSSIKIEVSTGV----- 43  
Db 24 VINNNKINETIGVSVTKPIYESVLOAKN-EKRSYARDLATYLVSEAEITHARVGHDAKR 82  
QY 44 -----TUTGNGQIS--IENTIPVS-----NENAGLITSPGAILLEASF 81  
Db 83 LHGDVDRDILTALATVGEVKAITRNEIAVHEHLGEYTNARFNDHAGLIL-NPRA--LDLRF 139  
QY 82 FINIISLPDISINVKEIQHQV-----VLTSGKSEITLKGKDVDPQYPRLOEV 129  
Db 140 FFSHWASIFALTEETARGIRHSIQFDDHGDLSLHKVYITDTHMDAWNLTLDITY-----L 194  
QY 130 STENPL--ILKTKLLKSIITATAFAASLQESRPLITGVHIV-----LSNHKDFKAVA 179  
Db 195 SPENPVLEITPAKSFDPATVTTALAQQLEQQWRSMTDVHQPFKILQENNLSPQAFKAVS 254  
QY 180 TDSHRMSQRLITDNLSADLMVPLPSKSLREFSAVETDDIETVEVFPSPQILFRSEHS 239  
Db 255 DD-----LAYQVONSALKTLIAL-AKEVQNEIMIFVGNRGCVQIFTG----- 295  
QY 240 FYTRLLEGNYPDTRLLMTEFETEV-----VENTQSLRHAME---RAFLISNATQGTGTV 290  
Db 296 -----KIDRLVPHQFENSEQVWINIFNPAPFTLHLIESEIVESWITRKPTQDGFV 344  
QY 291 -KLEI 294  
Db 345 TSLEL 349

Search completed: January 28, 2005, 17:01:08  
Job time : 44 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 16:47:03 ; Search time 71 Seconds  
(without alignments)  
3063.262 Million cell updates/sec

Title: US-10-048-071-28  
Perfect score: 1863  
Sequence: 1 MIQSFNRLTPIHALNTTKR.....LTPGDEBSFIQLITPVRTN 378

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_02.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1863	100.0	378	Q9EVR1	Q9evr1 streptococc
2	1849	99.2	378	Q9A209	Q9a209 streptococc
3	1835	98.5	378	Q8P329	Q8p329 streptococc
4	1832	98.3	378	Q79YT3	Q79yt3 streptococc
5	1832	98.3	378	Q8K829	Q8k829 streptococc
6	1527	82.0	378	Q8DMN8	Q8dmn8 streptococc
7	1491	80.0	378	Q8E2I6	Q8e2i6 streptococc
8	1491	80.0	378	Q8E7Z3	Q8e7z3 streptococc
9	1389	74.6	378	DP3B_STRPN	O06672 streptococc
10	1384	74.3	378	DP3B_STRR6	P59651 streptococc
11	1053	56.5	380	DP3B_LACLA	O9cjl1 lactococcus
12	1052	56.5	380	DP3B_LACLC	O54376 lactococcus
13	916	45.2	189	Q9L571	Q9l571 streptococc
14	864.5	46.4	376	Q839Z4	Q839z4 enterococcu
15	830	44.6	379	Q890K7	Q890k7 lactobacill
16	770.5	41.4	376	Q74M33	Q74m33 lactobacill
17	770.5	41.4	376	AA507982	AA507982 lactobaci
18	748.5	40.2	377	DP3B_STAEP	Q8cqe6 staphylococ
19	730.5	39.2	377	DP3B_STAAM	P50029 staphylococ
20	730.5	39.2	377	Q6GD88	Q6gd88 staphylococ
21	730.5	39.2	377	Q6GKU3	Q6gku3 staphylococ
22	722.5	38.8	378	Q8EU87	Q8eu87 oceanobacil
23	710	38.1	381	Q92FV1	Q92fv1 listeria in
24	703	37.7	381	Q8YAW1	Q8yaw1 listeria mo
25	703	37.7	381	Q725G9	Q725g9 listeria mo
26	703	37.7	381	AA702792	AA702792 listeria
27	695	37.3	379	Q6HQ02	Q6hq02 bacillus th
28	693	37.2	379	Q73FK4	Q73fk4 bacillus ce
29	693	37.2	379	AA538938	AA538938 bacillus
30	692	37.1	379	Q81W34	Q81w34 bacillus an
31	692	37.1	379	AA729080	AA729080 bacillus

RESULT 1

ID	Q9EVR1	PRELIMINARY;	PRT;	378 AA.
AC	Q9EVR1;			
DT	01-MAR-2001 (Tremblrel. 16, Created)			
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)			
DT	01-OCT-2003 (Tremblrel. 25, Last annotation update)			
DE	DNA polymerase III beta subunit (Fragment).			
GN	Name=dnaN;			
OS	Streptococcus pyogenes;			
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_TaxID=1314;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Bruck I., O'Donnell M.;			
RL	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF280765; AAF98349.2; --			
DR	GO; GO:0003677; F:DNA binding; IEA.			
DR	GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.			
DR	GO; GO:0006260; P:DNA replication; IEA.			
DR	InterPro; IPR001001; DNA_poliIII_beta.			
DR	Pfam; PF00712; DNA_poli3_beta_1.			
DR	Pfam; PF02767; DNA_poli3_beta_2; 1.			
DR	Pfam; PF02768; DNA_poli3_beta_3; 1.			
DR	SMART; SM00480; POL3bc; 1.			
DR	TIGRFAMs; TIGR00663; dnan; 1.			
FT	NON_TER 378 378			
SQ	SEQUENCE 378 AA; 41866 MW; 4663FP0B524C3D64 CRC64;			
Query Match	100.0%; Score 1863; DB 2; Length 378;			
Best Local Similarity	100.0%; Pred. No. 2.5e-115;			
Matches 378;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1 MIQSFNRLTPIHALNTTKRAISTKNAIPILUSSKIEVTSTGVTLTGNGQISIENTIPV 60			P05649 bacillus su
Db	1 MIQSFNRLTPIHALNTTKRAISTKNAIPILUSSKIEVTSTGVTLTGNGQISIENTIPV 60			Q81jd4 bacillus ce
Qy	61 SNEAGLLITSPGAILLEASFPINTISLDPISINVKIEHQVVLTSKSEITLKGKDV 120			Q8gqr3 bacillus we
Db	61 SNEAGLLITSPGAILLEASFPINTISLDPISINVKIEHQVVLTSKSEITLKGKDV 120			Q9rcal bacillus th
Qy	121 DQYPRLOEVSTENPLILTKLLKSIITAEFAASLQESRPILTGTVHIVLSNHKDFKAVAT 180			Q6bi40 bacillus an
Db	121 DQYPRLOEVSTENPLILTKLLKSIITAEFAASLQESRPILTGTVHIVLSNHKDFKAVAT 180			Q81pw0 bacillus an
Qy	181 DSHRMSQRLITLNTSADLMVLPKSLRFSASFVTTDDIETVEVFFSPQILFRSEHISF 240			Aat31800 bacillus
Db	181 DSHRMSQRLITLNTSADLMVLPKSLRFSASFVTTDDIETVEVFFSPQILFRSEHISF 240			Q81cqc4 bacillus ce
Qy	241 YTRLLEGNYPTDRLMTFETFEVFNFTQSHRAMEAFILSNATONGVKLEITONHIS 300			Aas41632 bacillus
Db	241 YTRLLEGNYPTDRLMTFETFEVFNFTQSHRAMEAFILSNATONGVKLEITONHIS 300			Q8rdl5 thermoaner
				Q97n34 clostridium
				Q8xpg1 clostridium
				Q6yrl2 onion yello

ALIGNMENTS

```

Qy 301 AHVNSPEVGKVEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFLSPVRPFTLT 360
Db 301 AHVNSPEVGKVEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFLSPVRPFTLT 360
Qy 361 PGDEESFIQLITPVRTN 378
Db 361 PGDEESFIQLITPVRTN 378

RESULT 2
Q9A209 Q9A209 PRELIMINARY; PRT; 378 AA.
AC Q9A209;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Beta subunit of DNA polymerase III (EC 2.7.7.7).
GN Name=dnaN; OrderedLocusNames=SPY0003;
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAINS=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071559398;
RA Ferretti J.C., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Prineaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AEO06472; AAK33147.1; -.
DR GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006260; F:DNA replication activity; IEA.
DR InterPro; IPR001001; DNA_polIII_beta.
DR Pfam; PF00712; DNA_pol3_beta_1.
DR Pfam; PF02767; DNA_pol3_beta_2; 1.
DR Pfam; PF02768; DNA_pol3_beta_3; 1.
DR SMART; SM00480; POL3BC; 1.
DR TIGRFAMs; TIGR00663; dnan; 1.
DR Complete proteome; Nucleotidyltransferase; Transferase.
KW SEQUENCE 378 AA; 41884 MW; 31DFF0FF326613D8 CRC64;
SQ SEQUENCE 378 AA; 41884 MW; 31DFF0FF326613D8 CRC64;

Query Match 99.2%; Score 1849; DB 2; Length 378;
Best Local Similarity 99.2%; Pred. No. 2.1e-114;
Matches 375; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MIQFSINTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGNGQISIENTIPV 60
Db 1 MIQFSINTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGNGQISIENTIPV 60
Qy 61 SNEAGLLITSPGAILLEASFFINISSLDPDISINVKIEHQVVLTSKGEITLKGKDV 120
Db 61 SNEAGLLITSPGAILLEASFFINISSLDPDISINVKIEHQVVLTSKGEITLKGKDV 120
Qy 121 DOYPLQEVSTENPLILTKLLKSIIAETAFASLQESRPILTGTVHIVLSNHKDFKAVAT 180
Db 121 DOYPLQEVSTENPLILTKLLKSIIAETAFASLQESRPILTGTVHIVLSNHKDFKAVAT 180
Qy 181 DSHRMSQRLITLDTNSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240
Db 181 DSHRMSQRLITLDTNSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240
Qy 241 YTRLLEGNYPTDRLMLTTEFEVFNQSLRHAMERAFILSNATQNGTVKLEITQNHIS 300
Db 241 YTRLLEGNYPTDRLMLTTEFEVFNQSLRHAMERAFILSNATQNGTVKLEITQNHIS 300
Qy 301 AHVNSPEVGKVEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFLSPVRPFTLT 360

```

```

Db 301 AHVNSPEVGKVEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFLSPVRPFTLT 360
Qy 361 PGDEESFIQLITPVRTN 378
Db 361 PGDEESFIQLITPVRTN 378

RESULT 3
Q8P329 Q8P329 PRELIMINARY; PRT; 378 AA.
AC Q8P329;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Beta subunit of DNA polymerase III.
GN Name=dnaN; OrderedLocusNames=spyM18_0002;
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=186103;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS8232; PubMed=11917108;
RX MEDLINE=21927593; PubMed=11917108;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Bares S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
DR EMBL; AEO09553; AAL96838.1; -.
DR GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0006260; F:DNA replication; IEA.
DR InterPro; IPR001001; DNA_polIII_beta.
DR Pfam; PF00712; DNA_pol3_beta_1.
DR Pfam; PF02767; DNA_pol3_beta_2; 1.
DR Pfam; PF02768; DNA_pol3_beta_3; 1.
DR SMART; SM00480; POL3BC; 1.
DR TIGRFAMs; TIGR00663; dnan; 1.
DR Complete proteome.
KW SEQUENCE 378 AA; 41863 MW; D2DB1C0CEFF16FD3 CRC64;
SQ SEQUENCE 378 AA; 41863 MW; D2DB1C0CEFF16FD3 CRC64;

Query Match 98.5%; Score 1835; DB 2; Length 378;
Best Local Similarity 98.4%; Pred. No. 1.8e-113;
Matches 372; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MIQFSINTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGNGQISIENTIPV 60
Db 1 MIQFSINTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGNGQISIENTIPV 60
Qy 61 SNEAGLLITSPGAILLEASFFINISSLDPDISINVKIEHQVVLTSKGEITLKGKDV 120
Db 61 SNEAGLLITSPGAILLEASFFINISSLDPDISINVKIEHQVVLTSKGEITLKGKDV 120
Qy 121 DOYPLQEVSTENPLILTKLLKSIIAETAFASLQESRPILTGTVHIVLSNHKDFKAVAT 180
Db 121 DOYPLQEVSTENPLILTKLLKSIIAETAFASLQESRPILTGTVHIVLSNHKDFKAVAT 180
Qy 181 DSHRMSQRLITLDTNSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240
Db 181 DSHRMSQRLITLDTNSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240
Qy 241 YTRLLEGNYPTDRLMLTTEFEVFNQSLRHAMERAFILSNATQNGTVKLEITQNHIS 300
Db 241 YTRLLEGNYPTDRLMLTTEFEVFNQSLRHAMERAFILSNATQNGTVKLEITQNHIS 300
Qy 301 AHVNSPEVGKVEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFLSPVRPFTLT 360

```

```
Db 301 AHVNSPEVKVKNEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFSLSPVPPFTLT 360
QY 361 PGDEESFIQLTPVRTN 378
Db 361 PGDEESFIQLTPVRTN 378

RESULT 4
Q79YT3
ID Q79YT3 PRELIMINARY; PRT; 378 AA.
AC Q79YT3;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DE Putative DNA polymerase III beta subunit.
GN OrderedLocusNames=Spa0002;
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_taxid=198466;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=SSI-1;
RX MEDLINE=22683278; PubMed=12799345;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution.";
RL Genome Res. 13:1042-1055(2003).
DR EMBL; AF005141; BAC63097.1; -
DR InterPro; IPR001001; DNA_polIII_beta.
DR Pfam; PF00712; DNA_pol3_beta_1.
DR Pfam; PF02767; DNA_pol3_beta_2; 1.
DR Pfam; PF02768; DNA_pol3_beta_3; 1.
DR SMART; SM00480; POL3bc; 1.
DR TIGRFAMs; TIGR00663; dnan; 1.
DR SEQUENCE 378 AA; 41861 MW; 24B19F5944A0E3B4 CRC64;
SQ SEQUENCE 378 AA; 41861 MW; 24B19F5944A0E3B4 CRC64;

Query Match 98.3%; Score 1832; DB 2; Length 378;
Best Local Similarity 98.1%; Pred. No. 2.9e-113;
Matches 371; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MIQPSINRTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGNGQISIENTIPV 60
Db 1 MIQPSINRTLFIHALNTTKRAISTKNAIPILSSIKIEVTPTGVTLTGNGQISIENTIPV 60
QY 61 SNEAGLLITSPGAILLEASFFINISSLPDISINVKIEHQVVLTSKGSEITLKGKDV 120
Db 61 SNEAGLLITSPGAILLEASFFINISSLPDININVKIEHQVVLTSKGSEITLKGKDV 120
QY 121 DOYPRLOEVSTENPLILTKLLKSIITAFAASIQESRPILTGVHIVLSNHKDFKAVAT 180
Db 121 DOYPRLOEVSTENPLILTKLLKSIITAFAASIQESRPILTGVHIVLSNHKDFKAVAT 180
QY 181 DSHRMSQRLITLNDTSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240
Db 181 DSHRMSQRLITLNDTSADFDVWIPCKSLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240
QY 241 YTRLLEGNYPTDRLMTTEFEVFNVTQSLRHAMERAFILSNATQNTGVKLEITQNHIS 300
Db 241 YTRLLEGNYPTDRLMTTEFEVFNVTQSLRHAMERAFILSNATQNTGVKLEITQNHIS 300
QY 301 AHVNSPEVKVKNEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFSLSPVPPFTLT 360
Db 301 AHVNSPEVKVKNEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFSLSPVPPFTLT 360
QY 361 PGDEESFIQLTPVRTN 378.
Db 361 PGDEESFIQLTPVRTN 378
```

RESULT 6

```
RESULT 5
Q8X8Z9
ID Q8X8Z9 PRELIMINARY; PRT; 378 AA.
AC Q8X8Z9;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Beta subunit of DNA polymerase III.
GN Name=dnaN; OrderedLocusNames=SpyM3_0002;
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_taxid=198466;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206;
RA Beres S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Perkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
DR EMBL; AE014136; AAM78609.1; -
DR GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR InterPro; IPR001001; DNA_polIII_beta.
DR Pfam; PF00712; DNA_pol3_beta_1.
DR Pfam; PF02767; DNA_pol3_beta_2; 1.
DR Pfam; PF02768; DNA_pol3_beta_3; 1.
DR SMART; SM00480; POL3bc; 1.
DR TIGRFAMs; TIGR00663; dnan; 1.
DR Complete proteome.
KW SEQUENCE 378 AA; 41861 MW; 24B19F5944A0E3B4 CRC64;
SQ SEQUENCE 378 AA; 41861 MW; 24B19F5944A0E3B4 CRC64;

Query Match 98.3%; Score 1832; DB 2; Length 378;
Best Local Similarity 98.1%; Pred. No. 2.9e-113;
Matches 371; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MIQPSINRTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGNGQISIENTIPV 60
Db 1 MIQPSINRTLFIHALNTTKRAISTKNAIPILSSIKIEVTPTGVTLTGNGQISIENTIPV 60
QY 61 SNEAGLLITSPGAILLEASFFINISSLPDISINVKIEHQVVLTSKGSEITLKGKDV 120
Db 61 SNEAGLLITSPGAILLEASFFINISSLPDININVKIEHQVVLTSKGSEITLKGKDV 120
QY 121 DOYPRLOEVSTENPLILTKLLKSIITAFAASIQESRPILTGVHIVLSNHKDFKAVAT 180
Db 121 DOYPRLOEVSTENPLILTKLLKSIITAFAASIQESRPILTGVHIVLSNHKDFKAVAT 180
QY 181 DSHRMSQRLITLNDTSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240
Db 181 DSHRMSQRLITLNDTSADFDVWIPCKSLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240
QY 241 YTRLLEGNYPTDRLMTTEFEVFNVTQSLRHAMERAFILSNATQNTGVKLEITQNHIS 300
Db 241 YTRLLEGNYPTDRLMTTEFEVFNVTQSLRHAMERAFILSNATQNTGVKLEITQNHIS 300
QY 301 AHVNSPEVKVKNEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFSLSPVPPFTLT 360
Db 301 AHVNSPEVKVKNEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFSLSPVPPFTLT 360
QY 361 PGDEESFIQLTPVRTN 378
Db 361 PGDEESFIQLTPVRTN 378
```

```

Q8DWN8      PRELIMINARY;      PRT;      378 AA.
ID Q8DWN8
AC Q8E216;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Putative DNA polymerase III, beta subunit (SC 2.7.7.7).
GN Name=dnaN; OrderedLocusNames=SMU.02;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=11309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UA159 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).
DR EMBL; AE014853; AAN57795.1; -.
DR GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR InterPro; IPR001001; DNA_polIII_beta.
DR Pfam; PF00712; DNA_pol3_beta_1.
DR Pfam; PF02767; DNA_pol3_beta_2; 1.
DR Pfam; PF02768; DNA_pol3_beta_3; 1.
DR SMART; SM00480; POL3bc_1.
DR TIGRFAMs; TIGR00663; dnan; 1.
DR Complete proteome; Nucleotidyltransferase; Transferase.
KW SEQUENCE 378 AA; 42074 MW; DP9822F73D31CAC6 CRC64;
SQ
Query Match      82.0%; Score 1527; DB 2; Length 378;
Best Local Similarity 79.9%; Pred. No. 4.7e-93;
Matches 302; Conservative 39; Mismatches 37; Indels 0; Gaps 0;
QY 1 MIQSFINTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGQISIENTIPV 60
DB 1 MIKFSINKVFFLQALNATKRAISSKNAIPILSSLKIEVNSQSITLTGSGQISIENTISA 60
QY 61 SNEAGLLITSPGAILLEASFFINISSLPDISINVKIEIQHVLTSGKSEITLKGKDV 120
DB 61 BEENAGLVTSGLALLEANFFINIVSSLPDITLDFEIEQHVLTSGKSEITLKGKDV 120
QY 121 DQYPRLOEVSTENPLILTKLLKSIATAFAAASLQESRPILTGVHIVLSNHNKDFKAVAT 180
DB 121 EQYPRLOEVGTNNPLILTKLLKTIISTATAAQSERSPILTGVLVLTNHNKDFKAVAT 180
QY 181 DSHRMSORLITLDNTSADLMVLPKSLRFSFAVPTDDIETVEVFFSPQILFRSEHISF 240
DB 181 DSHRMSORLITLDHSSDDFDVWIPSRSLRFAAFTDDIESVEVFFSPQILFRSEYISF 240
QY 241 YTRLLEGNYPTDRLMLTETETEVVNTQSLRHAMERAFILSNATQGTVKLEITQNHIS 300
DB 241 YTRLLEGNYPTDRLGLNFFETEVVFNALRAMEAHLISNATQGTVKLEITINNQT 300
QY 301 AHVNSPEVGKVNEDIDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFILSPVRPFLT 360
DB 301 AHVNSPEVGKVNEDIDIESLGNLITISFNPTYLIEALKALKSETVTIRFISPIRPFLT 360
QY 361 PGDEESFIQILTPVRTN 378
DB 361 PSDNSENFIQILTPVRTN 378

```

RESULT 7  
Q8E216

```

ID Q8E216      PRELIMINARY;      PRT;      378 AA.
AC Q8E216;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE DNA polymerase III, beta subunit.
GN Name=dnaN; OrderedLocusNames=SAG0002;
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=2222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
RA Tettelin H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S.N.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).
DR EMBL; AE014191; AAM98910.1; -.
DR TIGR; SAG0002; -.
DR GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR InterPro; IPR001001; DNA_polIII_beta.
DR Pfam; PF00712; DNA_pol3_beta_1.
DR Pfam; PF02767; DNA_pol3_beta_2; 1.
DR Pfam; PF02768; DNA_pol3_beta_3; 1.
DR TIGRFAMs; TIGR00663; dnan; 1.
DR Complete proteome.
KW SEQUENCE 378 AA; 42306 MW; 270D22643835C3AF CRC64;
SQ
Query Match      80.0%; Score 1491; DB 2; Length 378;
Best Local Similarity 77.5%; Pred. No. 1.1e-90;
Matches 293; Conservative 39; Mismatches 46; Indels 0; Gaps 0;
QY 1 MIQSFINTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGQISIENTIPV 60
DB 1 MIHESINKNFFLHALTVTKRAISHKNAIPILSTVKIEVTRDAITLTGSGQISIENTIPA 60
QY 61 SNEAGLLITSPGAILLEASFFINISSLPDISINVKIEIQHVLTSGKSEITLKGKDV 120
DB 61 SNEAGLLVTPGSIILEAGFFINISSLPDVTLETFEIEQHVLTSGKSEITLKGKDV 120
QY 121 DQYPRLOEVSTENPLILTKLLKSIATAFAAASLQESRPILTGVHIVLSNHNKDFKAVAT 180
DB 121 DQYPRLOEQMTDPTLTLETLLKLSIINETAFAASQESRPILTGVLVLSQNKYFKAVAT 180
QY 181 DSHRMSORLITLDNTSADLMVLPKSLRFSFAVPTDDIETVEVFFSPQILFRSEHISF 240
DB 181 DSHRMSORTFOLESANNFDLVFESKSLRFSFAVPTDDIETVEVFFSDSQMLFRSEHISF 240
QY 241 YTRLLEGNYPTDRLMLTETETEVVNTQSLRHAMERAFILSNATQGTVKLEITQNHIS 300
DB 241 YTRLLEGNYPTDRLTNQPTETIIFNTALRHAMERAYLISNATQGTVRLEIQNETVS 300
QY 301 AHVNSPEVGKVNEDIDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFILSPVRPFLT 360
DB 301 AHVNSPEVGKVNEDLTVSLKGSINISFNPTYLIESLKAVKSETVTIRFISPIRPFLT 360
QY 361 PGDEESFIQILTPVRTN 378
DB 361 PGDETFIQILTPVRTN 378

```

## RESULT 8

```

Q8E723      PRELIMINARY;          PRT;    378 AA.
ID  Q8E723
AC  Q8E723
DT  01-MAR-2003 (TREMELrel. 23, Created)
DT  01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT  01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE  Beta subunit of DNA polymerase III.
GN  Name=dnaN; OrderedLocNames=gbs0002;
OS  Streptococcus agalactiae (serotype III).
OC  Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC  Streptococcus.
OX  NCBI_taxID=216495;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=NEM316 / Serotype III;
RX  MEDLINE=22242508; PubMed=12354221;
RA  Glaeser P., Ruanio C., Buchrieser C., Chevalier F., Frangeul L.,
RA  Msadek T., Zouine M., Couve E., Lelioui L., Poyart C., Trieu-Cuot P.,
RA  Kunst F.;
RT  "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT  invasive neonatal disease."
RL  Mol. Microbiol. 45:1499-1513 (2002).
DR  EMBL; AL766843; CAD45647.1; -.
DR  SAGAList; gbs0002; -.
DR  GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR  GO; GO:0003677; F:DNA binding; IEA.
DR  GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.
DR  GO; GO:0006260; P:DNA replication; IEA.
DR  InterPro; IPR001001; DNA_polIII_beta.
DR  Pfam; PF00712; DNA_pol3_beta; 1.
DR  Pfam; PF02767; DNA_pol3_beta_2; 1.
DR  Pfam; PF02768; DNA_pol3_beta_3; 1.
DR  SMART; SM00480; POL3bc; 1.
DR  TIGRFAMs; TIGR00663; dnan; 1.
KW  Complete proteome.
SQ  SEQUENCE 378 AA; 42306 MW; 270D22643835C3AF CRC64;

```

## RESULT 9

```

DP3B_STRPN  STANDARD;          PRT;    378 AA.
ID  DP3B_STRPN
AC  O06672?
DT  15-DEC-1998 (Rel. 37, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  01-OCT-2004 (Rel. 45, Last annotation update)
DE  DNA polymerase III, beta chain (SC 2.7.7.7).
GN  Name=dnaN; OrderedLocNames=SP0002;
OS  Streptococcus pneumoniae.
OC  Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC  Streptococcus.
OX  NCBI_taxID=1313;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=R801;
RA  Gasc A.A.;
RL  Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=ATCC BAA-334 / TIGR4;
RX  MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RA  Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA  Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,
RA  Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA  Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA  Holtzapple E.K., Khouri H.M., Wolf A.M., Uterback T.R., Hansen C.L.,
RA  McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
RA  Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA  Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT  "Complete genome sequence of a virulent isolate of Streptococcus
RT  pneumoniae."
RL  Science 293:498-506 (2001).
CC  -!- FUNCTION: DNA polymerase III is a complex, multichain enzyme
CC  responsible for most of the replicative synthesis in bacteria.
CC  This DNA polymerase also exhibits 3' to 5' exonuclease activity.
CC  The beta chain is required for initiation of replication once it
CC  is clamped onto DNA, it slides freely (bidirectional and ATP-
CC  independent) along duplex DNA (By similarity).
CC  -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC  + {DNA} (N).
CC  -!- SUBUNIT: DNA polymerase III contains a core (composed of alpha,
CC  epsilon and theta chains) that associates with a tau subunit. This
CC  core dimerizes to form the POLIII' complex. POLIII' associates
CC  with the gamma complex (composed of gamma, delta, delta', psi and
CC  chi chains) and with the beta chain to form the complete DNA
CC  polymerase III complex (By similarity).
CC  -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF000658; AAC45337.1; -.
DR  EMBL; AE007318; AAK74195.1; -.
DR  PIR; B95000; B95000.
DR  PIR; B97872; B97872.
DR  TIGR; SP0002; -.
DR  InterPro; IPR001001; DNA_polIII_beta.
DR  Pfam; PF00712; DNA_pol3_beta; 1.
DR  Pfam; PF02767; DNA_pol3_beta_2; 1.
DR  Pfam; PF02768; DNA_pol3_beta_3; 1.
DR  SMART; SM00480; POL3bc; 1.
DR  TIGRFAMs; TIGR00663; dnan; 1.
KW  Complete proteome; DNA replication; DNA-directed DNA polymerase;
KW  Transferrase.
FT  CONFLICT 17 17 T -> I (in Ref. 1).
FT  CONFLICT 43 43 I -> V (in Ref. 1).
SQ  SEQUENCE 378 AA; 42072 MW; 5749CB6E8E57FFE6 CRC64;

```

```
Query Match 74.6%; Score 1389; DB 1; Length 378;
Best Local Similarity 72.2%; Pred. No. 6.6e-84;
Matches 273; Conservative 50; Mismatches 55; Indels 0; Gaps 0;

CC 1 MIQSFINTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTSGNGQISIENTIPV 60
CC 1 MIHFSINKNLFQALNTTKRAISSKNAIPILSTVKIDVTNEGTLTSGNGQISIENTFSQ 60
CC 61 SNEAGLLITSPGAILLEASFFINIISLDPISINVKIEHOHVLTSGKSEITLKGKDV 120
CC 61 KNEADAGLLITSLGILLLEASFFINVSLSLDPVTLDFKEIQNQIVLTSGKSEITLKGKDS 120
CC 121 DQYRLQEVSTENPLIKTLKLSIIAETAFAASLQESRPILTGVHIVLSNHKDFKAVAT 180
CC 121 EQYPRIQEISASTPLILETKLLKIIINETAFAASTQESRPILTVGHVLSOHKELKTAT 180
CC 181 DSHRMSQRLITLNTSADLMVLPKSLRSESAVFTDDIETVEFFSPQILFRSEHISF 240
CC 181 DSHRLSQKLTLEKNSDDFDVVPISRLREFSAVFTDDIETVEIFFANQLFRSENISF 240
CC 241 YTRLLEGNYPTDRLMTPEFETEVFNTQSLRHAMERAFILSNATQNGTVKLEITQNHIS 300
CC 241 YTRLLEGNYPTDRLIPTDFNTITFNVNLRQSMERARLLSSATQNGTVKLEIKDGVVS 300
CC 301 AHVSPGVKNEDLDIVSQSGDLTISFNPTYLIESLKAISKSETVKIHFSPVRPFTLT 360
CC 301 AHVSPGVKNEDIDTQVTGEDLTISFNPTYLIDSLKALNSEKVTISFISAVRPFTLV 360
CC 361 PGDEESFIQILTPVRTN 378
CC 361 PADTDEDFMQLITPVRTN 378

RESULT 10
DP3B_STRR6
ID DP3B_STRR6 STANDARD; PRT; 378 AA.
AC P59651;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE DNA polymerase III, beta chain (EC 2.7.7.7).
GN Name=dnaN; OrderedLocusNames=ep0002;
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RX DOI=10.1128/JB.183.19.5709-5717.2001;
RA Hoakins J., Alborn W.E. Jr., Arnold J., Blazczak L.C., Burgett S.,
RA Dehoff B.S., Estrem S.T., Fricz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RA "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
CC -!- FUNCTION: DNA polymerase III is a complex, multichain enzyme
CC responsible for most of the replicative synthesis in bacteria.
CC This DNA polymerase also exhibits 3' to 5' exonuclease activity.
CC The beta chain is required for initiation of replication once it
CC is clamped onto DNA, it slides freely (bidirectional and ATP-
CC independent) along duplex DNA (By similarity).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + (DNA)(N).
CC -!- SUBUNIT: DNA polymerase III contains a core (composed of alpha,
CC epsilon and theta chains) that associates with a tau subunit. This
CC core dimerizes to form the POLIII' complex. POLIII' associates
CC with the gamma complex (composed of gamma, delta, delta', psi and
```

```
chi chains) and with the beta chain to form the complete DNA
polymerase III complex (By similarity).
-!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AE008385; AAK98806.1;
InterPro: IPR001001; DNA_polIII_beta.
Pfam; PF00712; DNA_pol3_beta_1;
Pfam; PF02767; DNA_pol3_beta_2; 1.
Pfam; PF02768; DNA_pol3_beta_3; 1.
SMART; SM00480; POL3Bc; 1.
TIGRFAMs; TIGR00663; dnan; 1.
KW Complete proteome; DNA replication; DNA-directed DNA polymerase;
KW Transference; 378 AA; 42070 MW; 14CE29278184F906 CRC64;
SQ SEQUENCE 378 AA; 42070 MW; 14CE29278184F906 CRC64;
Query Match 74.3%; Score 1384; DB 1; Length 378;
Best Local Similarity 72.2%; Pred. No. 1.4e-83;
Matches 273; Conservative 49; Mismatches 56; Indels 0; Gaps 0;

QY 1 MIQSFINTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTSGNGQISIENTIPV 60
DB 1 MIHFSINKNLFQALNTTKRAISSKNAIPILSTVKIDVTNEGTLTSGNGQISIENTFSQ 60
QY 61 SNEAGLLITSPGAILLEASFFINIISLDPISINVKIEHOHVLTSGKSEITLKGKDV 120
DB 61 KNEADAGLLITSLGILLLEASFFINVSLSLDPVTLDFKEIQNQIVLTSGKSEITLKGKDS 120
QY 121 DQYRLQEVSTENPLIKTLKLSIIAETAFAASLQESRPILTGVHIVLSNHKDFKAVAT 180
DB 121 EQYPRIQEISASTPLILETKLLKIIINETAFAASTQESRPILTVGHVLSOHKELKTAT 180
QY 181 DSHRMSQRLITLNTSADLMVLPKSLRSESAVFTDDIETVEFFSPQILFRSEHISF 240
DB 181 DSHRLSQKLTLEKNSDDFDVVPISRLREFSAVFTDDIETVEIFFANQLFRSENISF 240
QY 241 YTRLLEGNYPTDRLMTPEFETEVFNTQSLRHAMERAFILSNATQNGTVKLEITQNHIS 300
DB 241 YTRLLEGNYPTDRLIPTDFNTITFNVNLRQSMERARLLSSATQNGTVKLEIKDGVVS 300
QY 301 AHVSPGVKNEDLDIVSQSGDLTISFNPTYLIESLKAISKSETVKIHFSPVRPFTLT 360
DB 301 AHVSPGVKNEDIDTQVTGEDLTISFNPTYLIDSLKALNSEKVTISFISAVRPFTLV 360
QY 361 PGDEESFIQILTPVRTN 378
DB 361 PADTDEDFMQLITPVRTN 378

RESULT 11
DP3B_LACLA
ID DP3B_LACLA STANDARD; PRT; 380 AA.
AC Q9CJL1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE DNA polymerase III, beta chain (EC 2.7.7.7).
GN Name=dnaN; OrderedLocusNames=LL0002;
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=11403;
RX MEDLINE=21235186; PubMed=11337471; DOI=10.1101/gr.169701;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
```



```

RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RL Lactis ssp. Lactis IL1403."
RL Genome Res. 11:731-753(2001).
CC -!- FUNCTION: DNA polymerase III is a complex, multichain enzyme
CC responsible for most of the replicative synthesis in bacteria.
CC This DNA polymerase also exhibits 3' to 5' exonuclease activity.
CC The beta chain is required for initiation of replication once it
CC is clamped onto DNA, it slides freely (bidirectional and ATP-
CC independent) along duplex DNA (By similarity).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA} (N).
CC
CC -!- SUBUNIT: DNA polymerase III contains a core (composed of alpha,
CC epsilon and theta chains) that associates with a tau subunit. This
CC core dimerizes to form the POLIII' complex. PolIII' associates
CC with the gamma complex (composed of gamma, delta, delta', psi and
CC chi chains) and with the beta chain to form the complete DNA
CC polymerase III complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE006239; AAK04100.1; -.
CC PIR; B86625; B86625.
CC InterPro; IPR001001; DNA_polIII_beta.
CC Pfam; PF00712; DNA_pol3_beta_1.
CC Pfam; PF02767; DNA_pol3_beta_2; 1.
CC Pfam; PF02768; DNA_pol3_beta_3; 1.
CC SMART; SM00480; POL3bc; 1.
CC TIGRFAMs; TIGR00663; dnan; 1.
CC Complete proteome; DNA replication; DNA-directed DNA polymerase;
KW Transferrase.
SQ SEQUENCE 380 AA; 42264 MW; 2D54BD83932029B0 CRC64;

Query Match 56.5%; Score 1053; DB 1; Length 380;
Best Local Similarity 54.7%; Pred. No. 1.2e-61;
Matches 208; Conservative 73; Mismatches 97; Indels 2; Gaps 1;

QY 1 MIQSFNTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGTGVLTSNGQISIENTIPV 60
DB 1 MIKESINKNAQNARITKQALGSKVTIPALTKLIEVEENGITLIGSNGQISIKNPLPV 60

QY 61 SNEAGLLITSPGAILLEASFPINIISSLPDISINVKIEHQHVLTSGKSEITLKGKDV 120
DB 61 DNKDSMLISGTGVLLEAAFPENVVSQLPEVLTFTKEQKQVLTSGKSEITLKGKLD 120

QY 121 DOYPLQEVSTENPLILKTLKLSIAETAFAASLQESRPILTGVHIVLSNHKDFKAVAT 180
DB 121 EIYPHLQISEGSSLKMKVKVLEIFETVFAVSTQENRPIFTGVHLETLSTGELKAVAT 180

QY 181 DSHRMSQRLITLDNTSADLMVLPSPKSLREFSAVETDDIETVEVFPSPQILFRSEHSF 240
DB 181 DSHRMSQRLITLDNTSADLMVLPSPKSLREFSAVETDDIETVEVFPSPQILFRSEHSF 240

QY 241 YTRLLEGNYPDTRLAMTE--FETEVFNQSLRHAMERAFILSNATQNGTVKLEITQNH 298
DB 241 YSRLLIEGSPDTRNLIPNEADYTLDLVFDAAQLRHTMDRRLITVMTTNGTVKLVSGDS 300

QY 299 ISAHVNSPEVGKWNEDLDIVSOGSDLIISFNPTYLIESLKAISKSETVKIHLSPVRPT 358
DB 299 ISAHVNSPEVGKWNEDLDIVSOGSDLIISFNPTYLIESLKAISKSETVKIHLSPVRPT 358

QY 301 VYTTANSPEVGSHVELTALSKEGNDLAI SFNPEYLLIDALKVKAPEVRIRFISNVRPT 360
DB 301 VYTTANSPEVGSHVELTALSKEGNDLAI SFNPEYLLIDALKVKAPEVRIRFISNVRPT 360

QY 359 LTPGDEERSFICLITPVRTN 378
DB 361 LQPRNEESGFVQLITPVRTN 380

```

```

RESULT 12
DP3B LACLC STANDARD; PRT; 380 AA.
AC 054376;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE DNA polymerase III, beta chain (EC 2.7.7.7).
GN Name=dnan;
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OX Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
CX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGI363;
RX MEDLINE=98118563; PubMed=9435243;
RA El-Karoui M., Ehrlich S.D., Gruss A.;
RT "Identification of the lactococcal exonuclease/recombinase and its
modulation by the putative Chi sequence."
RL Proc. Natl. Acad. Sci. U.S.A. 95:626-631(1998).
CC -!- FUNCTION: DNA polymerase III is a complex, multichain enzyme
CC responsible for most of the replicative synthesis in bacteria.
CC This DNA polymerase also exhibits 3' to 5' exonuclease activity.
CC The beta chain is required for initiation of replication once it
CC is clamped onto DNA, it slides freely (bidirectional and ATP-
CC independent) along duplex DNA (By similarity).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA} (N).
CC
CC -!- SUBUNIT: DNA polymerase III contains a core (composed of alpha,
CC epsilon and theta chains) that associates with a tau subunit. This
CC core dimerizes to form the POLIII' complex. PolIII' associates
CC with the gamma complex (composed of gamma, delta, delta', psi and
CC chi chains) and with the beta chain to form the complete DNA
CC polymerase III complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U76424; AAC12964.1; -.
CC PIR; T30306; T30306.
CC InterPro; IPR001001; DNA_polIII_beta.
CC Pfam; PF00712; DNA_pol3_beta_1.
CC Pfam; PF02767; DNA_pol3_beta_2; 1.
CC Pfam; PF02768; DNA_pol3_beta_3; 1.
CC SMART; SM00480; POL3bc; 1.
CC TIGRFAMs; TIGR00663; dnan; 1.
CC DNA replication; DNA-directed DNA polymerase; Transferrase.
KW SEQUENCE 380 AA; 42209 MW; 4877297F3E822644 CRC64;

Query Match 56.5%; Score 1052; DB 1; Length 380;
Best Local Similarity 54.7%; Pred. No. 1.4e-61;
Matches 208; Conservative 74; Mismatches 96; Indels 2; Gaps 1;

QY 1 MIQSFNTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGTGVLTSNGQISIENTIPV 60
DB 1 MIKESINKNAQNARITKQALGSKVTIPALTKLIEVEENGITLIGSNGQISIKNPLPA 60

QY 61 SNEAGLLITSPGAILLEASFPINIISSLPDISINVKIEHQHVLTSGKSEITLKGKDV 120
DB 61 DNKDSMLISGTGVLLEAAFPENVVSQLPEVLTFTKEQKQVLTSGKSEITLKGKLD 120

QY 121 DOYPLQEVSTENPLILKTLKLSIAETAFAASLQESRPILTGVHIVLSNHKDFKAVAT 180
DB 121 EIYPHLQISEGSSLKMKVKVLEIFETVFAVSTQENRPIFTGVHLETLSTGELKAVAT 180

QY 181 DSHRMSQRLITLDNTSADLMVLPSPKSLREFSAVETDDIETVEVFPSPQILFRSEHSF 240
DB 181 DSHRMSQRLITLDNTSADLMVLPSPKSLREFSAVETDDIETVEVFPSPQILFRSEHSF 240

```

```

Db 181 DSHRMSQRLPLEDTLKFDPVLPKSKINSFNKVTNDEEIEIFISQMLPQNETISY 240
Qy 241 YTRLLEGNYPTDRLLMTE--FETEVVENTOSLRHAMERAFNISNATONGTVKLEITONH 298
Db 241 YSRLIEGSPNRLNLPNEAYTLDLVEDAQLRHTWDRALLTWITNGTVKLVTSGDS 300
Qy 299 ISAHVNSPEVGVKNEDLDIVSQSGDLTISFNPTYLIESLKAISKSETVKIHLSPVRPPT 358
Db 301 VVTANSPEVGSVHEELTALSKEGNDLSISFNPEYLDALKVIRKAPVIRFISNVRPPT 360
Qy 359 LTPGDEESFQILTPVPTN 378
Db 361 LQPRNEESGFVLITPVPTN 380

RESULT 13
Q9L571 PRELIMINARY; PRT; 189 AA.
AC Q9L571;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE DNA polymerase III beta subunit (EC 2.7.7.7) (Fragment).
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370;
RX MEDLINE=20389620; PubMed=10930754;
RA Suvorova A.N., Ferrretti J.J.;
RT "Replication origin of streptococcus pyogenes, organization and
RL cloning in heterologous systems.";
RL FEMS Microbiol. Lett. 189:293-297(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370;
RA Suvorov A., Ferrretti J.J.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF255728; AAF71536.1; -.
DR GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR InterPro; IPR001001; DNA_polIII_beta.
DR Pfam; PF00712; DNA_pol3_beta_1.
DR Pfam; PF02767; DNA_pol3_beta_2; 1.
DR SMART; SM00480; POL3Bc; 1.
KW Nucleotidyltransferase; Transferase.
FT NON TER 189
SQ SEQUENCE 189 AA; 20570 MW; BDDC423D8AB1C441 CRC64;

Query Match 49.2%; Score 916; DB 2; Length 189;
Best Local Similarity 100.0%; Pred. No. 5.7e-53;
Matches 189; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIQPSINTLFHALNTTKRAISTKNAIPILSSIKIEVTSVTLTGSGQISIENTIPV 60
Db 1 MIQPSINTLFHALNTTKRAISTKNAIPILSSIKIEVTSVTLTGSGQISIENTIPV 60
Qy 61 SNEAGLLITSPGAILLEASFFINISSLPDISINVKIEHQHVLTSGKSEITLKGKDV 120
Db 61 SNEAGLLITSPGAILLEASFFINISSLPDISINVKIEHQHVLTSGKSEITLKGKDV 120
Qy 121 DQYPRLOQVSTENPLILTKLKSIITAETAFAASIQESRPILTGVIHVLNHNKDFKAVAT 180
Db 121 DQYPRLOQVSTENPLILTKLKSIITAETAFAASIQESRPILTGVIHVLNHNKDFKAVAT 180
Qy 181 DSHRMSQRL 189
Db 181 DSHRMSQRL 189

```

```

RESULT 14
Q83924 PRELIMINARY; PRT; 376 AA.
ID Q83924
AC Q83924;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DNA polymerase III, beta subunit.
GN Name=dnaN; OrderedLocusNames=EF0002;
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V583 / ATCC 700802;
RX MEDLINE=2250857; PubMed=12663927; DOI=10.1126/science.1080613;
RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seehadri R.,
RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
RA Tettelin H., Dodson R.J., Umayam L.A., Brinkac L.M., Beanan M.J.,
RA Dougherty S.C., DeBoy R.T., Durkin S.A., Kolonay J.F., Madupu R.,
RA Nelson W.C., Vamathevan J.J., Tran B., Upton J., Hansen T., Shetty J.,
RA Khouri H.M., Utterback T.R., Radune D., Ketchum K.A., Dougherty B.A.,
RA Fraser C.M.;
RT "Role of mobile DNA in the evolution of vancomycin-resistant
RT Enterococcus faecalis.";
RL Science 299:2071-2074(2003).
DR EMBL; AE016947; AAC79887.1; -.
DR TIGR; EF0002; -.
DR GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR InterPro; IPR001001; DNA_polIII_beta.
DR Pfam; PF00712; DNA_pol3_beta_1.
DR Pfam; PF02767; DNA_pol3_beta_2; 1.
DR Pfam; PF02768; DNA_pol3_beta_3; 1.
DR SMART; SM00480; POL3Bc; 1.
DR TIGRfams; TIGR00663; dnan; 1.
KW Complete proteome.
SQ SEQUENCE 376 AA; 42214 MW; 2D992530A04093B4A CRC64;

Query Match 46.4%; Score 864.5; DB 2; Length 376;
Best Local Similarity 45.9%; Pred. No. 3.8e-49;
Matches 173; Conservative 77; Mismatches 126; Indels 1; Gaps 1;

Qy 2 IQPSINTLFHALNTTKRAISTKNAIPILSSIKIEVTSVTLTGSGQISIENTIPVS 61
Db 1 MKLTVKRSVFLQELQTVQRAISSKTTIPILTGKIVLSEDLGSLTGSNADISIESFLSKD 60
Qy 62 NENAGLLITSPGAILLEASFFINISSLPDISINVKIEHQHVLTSGKSEITLKGKDV 121
Db 61 DEKAQMTIERTGSIVLQSRPFGEIIRKLPEDMFTVEVDNNQVAITSKADFTVNGLDAD 120
Qy 122 QYPRLOQVSTENPLILTKLKSIITAETAFAASIQESRPILTGVIHVLNHNKDFKAVAT 181
Db 121 NYPHLPVIDIQNMKLPVLLTKIISSETGFAVSMHESRPILTGVIHFILENQK-LLA VATD 179
Qy 182 SHRMSQRLITLDNTSADLMVLPKSLREFSAVTDIDIEVEVFPFSQIILFREHSISFY 241
Db 180 SHRLSORVITEQVAEDFNIVIPGKSLTSLRSITNEEEMVEISIMENQVLFKTTMYFY 239
Qy 242 TRLEGNVPDTRLLMTEFETEVVENTOSLRHAMERAFNISNATONGTVKLEITONHISA 301
Db 240 SRLLEGNVPDTRLLMTEFETEVVENTOSLRHAMERAFNISNATONGTVKLEITONHISA 299
Qy 302 HVNSPEVGVKNEDLDIVSQSGDLTISFNPTYLIESLKAISKSETVKIHLSPVRPPTTP 361
Db 300 YGNSPEIGKVEEALNENVSGEALDISFNPDYMKDALRAFGDMNITVKFLSPINPPTLEP 359
Qy 362 GDEESFIQLITPVPTN 378

```

```
Db 360 TETELDFIQLITPVRTN 376
RESULT 15
Q890K7 PRELIMINARY; PRT; 379 AA.
ID Q890K7
AC Q890K7
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE DNA-directed DNA polymerase III, beta chain (EC 2.7.7.7).
GN Name=dnaN; OrderedLocusNames=lp_0002;
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 8826 / WCFS1;
RX MEDLINE=2480296; PubMed=12566566;
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M.,
RA Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B.,
RA De Vos W.M., Siezen R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCFS1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
DR EMBL; AL935252; CAD62704.1;
DR GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006260; F:DNA replication; IEA.
DR InterPro; IPR001001; DNA polIII beta.
DR Pfam; PF00712; DNA_pol3_beta_2; 1.
DR Pfam; PF02767; DNA_pol3_beta_3; 1.
DR Pfam; PF02768; DNA_pol3_beta_3; 1.
DR SMART; SM00480; POL3bc; 1.
DR TIGRFAMs; TIGR00663; dnan; 1.
KW Complete proteome; DNA-directed DNA polymerase;
KW Nucleotidyltransferase; Transferase.
SQ SEQUENCE 379 AA; 41520 MW; 13EA997B81E73717 CRC64;

Query Match 44.6%; Score 830; DB 2; Length 379;
Best Local Similarity 45.9%; Pred. No. 7.4e-47;
Matches 174; Conservative 73; Mismatches 128; Indels 4; Gaps 3;

Qy 2 IQPSINRFLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNGOISIENTIPVS 61
Db 1 MKFTINRSAPFKELANVQRAISSKTTIPLTLGKLDVNTDAITLTGSDADISIIETIPAS 60

Qy 62 NENAGLLITSPGAILLEASFFNIITSSLPDISINVKIEHQVLTSGKSEITLKGKVD 121
Db 61 DNNNTLVVEDAGSIVLPARFFSEIVKVLPEDTMTNVVVDGFTQITSGAASFTINGLDP 120

Qy 122 QYPRLOEVSTENPLIKTKLKSIIAETAFASLOESRPLITGVHIVLSNHNKDFKAVATD 181
Db 121 NYPHLPEDITNTTITLAGDVLKELICQTVIAVSNQESRPLITGVHIFILAN-GEFLAVATD 179

Qy 182 SHRMSORLITL-DNTSADLMVLPKSLREFSAVTDDETVEVFSPQILFRSEHLSF 240
Db 180 SHRLSORRIKLPEANNANYDVIIIPGKSLTELGRMIGDNNPDVQMLSENQVLFVIGNTSF 239

Qy 241 YTRLLEGNYPDTRLMTETEFETVFNQSLRHAMERAFISNATQNGTVKLEI--TQNH 298
Db 240 YSRLLEGNYPDTRSLIPKESNTTVISAPALSAALERASLSHESRNVVRSVNPDKT 299

Qy 299 ISAHVNSPEVGVKNEDLDIVSQSGDLTISFNPTLYLSLKAIKSETVKIHLSPV 358
Db 300 ITIFGNSPDVGSVTEQLQPTDLSGDELEISFNPDYKKEALRSGQAMIKISFTMALRPFT 359

Qy 359 LTPGDEESFIQLITPVRT 377
| : : : : : |

Db 360 LVPTEGENFIQLITPVRT 378
RESULT 16
Q74M33 PRELIMINARY; PRT; 376 AA.
ID Q74M33
AC Q74M33
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE DNA polymerase III, beta chain.
GN OrderedLocusNames=Lj0002;
OS Lactobacillus johnsonii.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=33959;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 533;
RX PubMed=14966310;
RA Fridmore R.D., Berger B., Desiere F., Vilanova D., Barretto C.,
RA Pittet A.-C., Zwahlen M.-C., Rouvet M., Altermann E., Barrangou R.,
RA Mollet B., Mercenier A., Klaenhammer T., Arigoni F., Schell M.A.;
RT "The genome sequence of the probiotic intestinal bacterium
RT Lactobacillus johnsonii NCC 533.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:2512-2517(2004).
DR EMBL; AE017200; AAS07982.1;
DR InterPro; IPR001001; DNA polIII beta.
DR Pfam; PF00712; DNA_pol3_beta_1;
DR Pfam; PF02767; DNA_pol3_beta_2; 1.
DR Pfam; PF02768; DNA_pol3_beta_3; 1.
DR SMART; SM00480; POL3bc; 1.
DR TIGRFAMs; TIGR00663; dnan; 1.
KW Complete proteome.
SQ SEQUENCE 376 AA; 41819 MW; 33FD6403EC85AA28 CRC64;

Query Match 41.4%; Score 770.5; DB 2; Length 376;
Best Local Similarity 44.4%; Pred. No. 6.4e-43;
Matches 170; Conservative 73; Mismatches 125; Indels 15; Gaps 7;

Qy 2 IQPSINRFLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNGOISIENTIPVS 61
Db 1 MQFTINRNLFIENLNANRAISSRATIPILSGIKLTLTDEMILTLGSDTDSIEIQIPVN 60

Qy 62 NENAGLLITSPGAILLEASFFNIITSSLPDISINVKIEHQVLTSGKSEITLKGK 119
Db 61 DD--LIQVSTGSIVLPARFFSEIVKVLPGKDFSEVKESFQTKIV--SENTEFMINGLD 115

Qy 120 VDQYPRLOEVSTENPLIKTKLKSIIAETAFASLOESRPLITGVHIVLSNHNKDFKAVA 179
Db 116 ANNYPHLPETISDASFKISGKTFREIINETVFAVATQESRPTLTGVNFIF--NNSSIKAVA 174

Qy 180 TDSHRMSORLITLQND--TSADLMVLPKSLREFSAVTDDETVEVFSPQILFRSE 236
Db 175 TDSHRLSORQLSLENGPOTSTDL--IIPKSLVELSRIIGESDPEITVNPGENQVLFV 232

Qy 237 HISPYTRLLEGNYPDTRLMTETEFETVFNQSLRHAMERAFISNATQNGTVK--LEI 294
Db 233 NIAPYSRLDGOYDPTDRLIPTSTSTTSVEFELPVLARSLERASLLTHESRNVVQMTLDV 292

Qy 295 TQNHSAHVNSPEVGVKNEDLDIVSQSGDLTISFNPTLYLSLKAIKSETVKIHLSPV 354
Db 293 QNQLVKLGQSDPFGNVVEEIGPKNLEGDGLTISFNPDYLRALREALRSTDSIMNFTQPL 352

Qy 355 RPFTLTPGDEESFIQLITPVRT 377
| : : : : : |
Qy 353 RPFTVIPAKQDVNFQLITPVRT 375

RESULT 17
AAS07982 PRELIMINARY; PRT; 376 AA.
ID AAS07982
AC AAS07982;
```

DT 02-MAR-2004 (TrEMBLrel. 27, Created)  
 DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)  
 DE DNA polymerase III, beta chain.  
 GN L370002.  
 OS Lactobacillus johnsonii.  
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
 OC Lactobacillus.  
 NCBI\_TaxID=33959;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCC 533;  
 RX PubMed=14966310;  
 RA Fridmore R.D., Berger B., Desiere F., Vilanova D., Barretto C.,  
 RA Fittet A.-C., Zwahlen M.-C., Rouvet M., Altermann E., Barrangou R.,  
 RA Mollat B., Mercenier A., Klaenhammer T., Arigoni F., Schell M.A.,  
 RT "The genome sequence of the probiotic intestinal bacterium  
 RT Lactobacillus johnsonii NCC 533".  
 RL Proc. Natl. Acad. Sci. U.S.A. 101:2512-2517(2004).  
 DR EMBL: A5017200; AAS07982.1; --  
 SQ SEQUENCE 376 AA; 41819 MW; 33FD6403EC85AA28 CRC64;

Query Match 41.4%; Score 770.5; DB 2; Length 376;  
 Best Local Similarity 44.4%; Pred. No. 6.4e-43;  
 Matches 170; Conservative 73; Mismatches 125; Indels 15; Gaps 7;

QY 2 IQPSINRTLIHALNTRKRAISTKNAIPILSSIKIEVTSTGVLGSGNGQISIENTIPVS 61  
 DB 1 MOFTINRNLFLNANRAISRATIPILSGIKLTLDMLTGLSGSDTISIEIQIPVN 60  
 QY 62 NENAGLITSPGAILLEAFINIISSLP--DISINVKIEHQVLTSGSKSEITLKGKD 119  
 DB 61 DD---LIVOSTGIVLPARFFSEIVKVLPGKDFSEVKESPTKIV--SENTEPMINGLD 115  
 QY 120 VDQYPRLOEVSTENPLILTKLKSIIAETAFASLQESRPILTGVHIVLSNKHDKFAVA 179  
 DB 116 ANNYPHLPETSDASFKISGKTRFRIINETVFAVATQESRPTLTGVNFI--NNSIKAVA 174  
 QY 180 TDSHRMSORLITLDN---TSADLMVLPKSLRPSAVFTDDIETVVFSPQILFRSE 236  
 DB 175 TDSHRLSQRQISLENGPQTSTDL--IIPKSLVLSRIIGSDPEITVNPGENQVLPVVG 232  
 QY 237 HISFYTRLLEGNYPTDRLIMTEPETEVVENTQSLRHAMERAFISNATONGTVK--LEI 294  
 DB 233 NIAFYSRLDGGYPTDRLIPETSTVEFELPVLARSLERASLITHERNNVVKMTLDV 292  
 QY 295 TONHISAHVNSPEVGKVNEDLDIVSQSGDLTISFNPTYLIESLKAISKSETVKIHFILSPV 354  
 DB 293 QNQLVKLQDGSPEIGNVEEIGFKNLEGDGLTISFPDYLREALRASITDSDIINWFTQPL 352  
 QY 355 RPFLITPGDEESFQILITPVRT 377  
 DB 353 RPFTVIPAKQDWNFTQILITPVRT 375

RESULT 18  
 DP3B\_STAEF  
 ID DP3B\_STAEF STANDARD; PRT; 377 AA.  
 AC Q8CQK6;  
 DT 29-MAR-2004 (Rel. 43, Created)  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE DNA polymerase III, beta chain (EC 2.7.7.7).  
 GN Name=dnaN; OrderedLocuNames=SE00002;  
 OS Staphylococcus epidermidis.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 NCBI\_TaxID=1282;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 12228;  
 RX PubMed=12950322;  
 RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,  
 RA Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,

RA Yuan Z.-H., Zhao G.-P., Qu D., Danchin A., Wen Y.-M.;  
 RT "Genome-based analysis of virulence genes in a non-biofilm-forming  
 RT Staphylococcus epidermidis strain (ATCC 12228).";  
 RL Mol. Microbiol. 49:1577-1593(2003).  
 CC -!- FUNCTION: DNA polymerase III is a complex, multichain enzyme  
 CC responsible for most of the replicative synthesis in bacteria.  
 CC This DNA polymerase also exhibits 3' to 5' exonuclease activity.  
 CC The beta chain is required for initiation of replication once it  
 CC is clamped onto DNA. It slides freely (bidirectional and ATP-  
 CC independent) along duplex DNA (By similarity).  
 CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
 CC + {DNA}(N).  
 CC -!- SUBUNIT: DNA polymerase III contains a core (composed of alpha,  
 CC epsilon and theta chains) that associates with a tau subunit. This  
 CC core dimerizes to form the POLIII' complex. POLIII' associates  
 CC with the gamma complex (composed of gamma, delta, delta', psi and  
 CC chi chains) and with the beta chain to form the complete DNA  
 CC polymerase III complex (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL: A5016744; AAO03599.1; --  
 DR InterPro: IPR001001; DNA\_polIII\_beta.  
 DR Pfam: PF00712; DNA\_pol3\_beta; 1.  
 DR Pfam: PF02767; DNA\_pol3\_beta\_2; 1.  
 DR Pfam: PF02768; DNA\_pol3\_beta\_3; 1.  
 DR TIGRFAMs: TIGR00663; dnai1.  
 DR Complete proteome; DNA replication; DNA-directed DNA polymerase;  
 KW Transference.  
 SQ SEQUENCE 377 AA; 42101 MW; 4189F2AFDF3BCBDD CRC64;

Query Match 40.2%; Score 748.5; DB 1; Length 377;  
 Best Local Similarity 40.5%; Pred. No. 1.9e-41;  
 Matches 153; Conservative 86; Mismatches 136; Indels 3; Gaps 3;

QY 1 MIQPSINRTLIHALNTRKRAISTKNAIPILSSIKIEVTSTGVLGSGNGQISIENTIPV 60  
 DB 1 MBEFTIKRDYFINQLANDTLKAIISPTTLPITGKIDAKENEVILTSGSDSEISIEITPK 60  
 QY 61 SNENAGLL--ITSPGAILLEAFINIISSLPDISINVKIEHQVLTSGSKSEITLKGKD 119  
 DB 61 QVDGEEIVEITETGVSVPVLPGRFFVDIKKLPGKEVKLSTNEQPTLTITSGHSEFNLSGLD 120  
 QY 120 VDQYPRLOEVSTENPLILTKLKSIIAETAFASLQESRPILTGVHIVLSNKHDKFAVA 179  
 DB 121 PDQYPLLPVSRDDAIQLSVKVLKNIIAQTNFAVSTSETRPVLTVGNWLIQDN--ELICTA 179  
 QY 180 TDSHRMSORLITLDNTSADLMVLPKSLRPSAVFTDDIETVVFSPQILFRSEHIS 239  
 DB 180 TDSHRLAVRLKQLEDESENKNVPIPGKALSSELNKMISDSDDEDIDIFFASGNVLPFRVGNIN 239  
 QY 240 FYTRLLEGNYPTDRLIMTEPETEVVENTQSLRHAMERAFISNATONGTVKLEITQNH 299  
 DB 240 FISRLLEGHPDTRFLPPENYEIKLGINNGDFVHAIDRASLIAREGGNNVILKSTGNELV 299  
 QY 300 SAHVNPSPEVGKVNEDLDIVSQSGDLTISFNPTYLIESLKAISKSETVKIHFILSPVFPF 359  
 DB 300 ELSSTSPFETGVKEVNANDVEGNNLKISFNKSNYMDALKAINDEVEVEFFCTMPPFIL 359  
 QY 360 TPGDEESFQILITPVRT 377  
 DB 360 KPXD--DSDVTQLILPRT 376

RESULT 19  
 DP3B\_STAAM

```

ID AC DP3B STAAW STANDARD; PRT; 377 AA.
DT P50029;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE DNA polymerase III, beta chain (EC 2.7.7.7).
GN Name=dnaN; OrderedLocusNames=SAV0002, SA0002, MW0002;
OS Staphylococcus aureus (strain Mu50 / ATCC 700699);
OS Staphylococcus aureus (strain N315);
OS Staphylococcus aureus (strain MW2), and
OC Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879, 196620, 1280;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=MU50 / ATCC 700699, and N315;
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanchisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayaishi H., Hiramatsu K.;
RA "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=MW2;
RX MEDLINE=22040717; PubMed=12044378;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
RT acquired MRSA.";
RL Lancet 359:1819-1827(2002).
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=YB886;
RX MEDLINE=95206242; PubMed=7898435;
RA Alonso J.C., Fisher L.M.;
RT "Nucleotide sequence of the recF gene cluster from Staphylococcus
RT aureus and complementation analysis in Bacillus subtilis recF
RT mutants.";
RL Mol. Gen. Genet. 246:680-686(1995).
CC -!- FUNCTION: DNA polymerase III is a complex, multichain enzyme
CC responsible for most of the replicative synthesis in bacteria.
CC This DNA polymerase also exhibits 3' to 5' exonuclease activity.
CC The beta chain is required for initiation of replication once it
CC is clamped onto DNA, it slides freely (bidirectional and ATP-
CC independent) along duplex DNA (By similarity).
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA} (N).
CC -!- SUBUNIT: DNA polymerase III contains a core (composed of alpha,
CC epsilon and theta chains) that associates with a tau subunit. This
CC core dimerizes to form the PolIII' complex. PolIII' associates
CC with the gamma complex (composed of gamma, delta, delta', psi and
CC chi chains) and with the beta chain to form the complete DNA
CC polymerase III complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP003358; BAB56164.1; -
CC EMBL; AP003129; BAB41218.1; -
CC EMBL; AP004822; BAB93867.1; -

```

---

```

DR PIR, S54708; S54708
DR SWISS-2DPAGE; P99103; STAAW.
DR InterPro; IPR01001; DNA_polIII_beta.
DR Pfam; PF00712; DNA_pol3_beta_1.
DR Pfam; PF02767; DNA_pol3_beta_2; 1.
DR Pfam; PF02768; DNA_pol3_beta_3; 1.
DR SMART; SM00480; POL3BC; 1.
DR TIGRFAMs; TIGR00663; dnan; 1.
KW Complete proteome; DNA replication; DNA-directed DNA polymerase;
KW Transference.
SQ SEQUENCE 377 AA; 41913 MW; 0A985EF94E044FBC CRC64;

Query Match 39.2%; Score 730.5; DB 1; Length 377;
Best Local Similarity 39.3%; Pred. No. 2.9e-40;
Matches 149; Conservative 90; Mismatches 135; Indels 5; Gaps 4;

QY 1 MIOFSINRFLTHALNTTKRAISNTKNAIPILSSIKIEVTSVTGVLITGSGOISIENTIPV 60
DB 1 MMEFTIKRDYFITQLNDTLKAIISPRITLPIITGKIDAKEHEVILTGSSEISIEITIPK 60
QY 61 SNEAGLL-ITSPGAILLEASFFINISLSPDISINVKIEIQHQVLTSGKSEITLTKGD 119
DB 61 TVDGEDIVINSETGVSVPGRFFVDIIKKLPKDKVLTSTNEQFQTLITSGHSEFNLGLD 120
QY 120 VDOYPRLOEVSTENPLILKTKLSITAEATAFAASLOESRPILITGVHIVLSNHHKDFKAVA 179
DB 121 PDQYPLLPQVSRDDAIQLSVKLVKNVIAQTNFVSTSETRPVLTGVNNLIQEN-ELICTA 179
QY 180 TDSHRMSORLITLNDTSADLMVVLPSKSLRFSFAVFTDDIETVEVFFSPSQILFRSEHIS 239
DB 180 TDSHRLAVRKLQLEDVSENKNVILPGKALAEINKIMSDNEEDIDIFFASQVLFKVGNNV 239
QY 240 FYTLLEGNYPTDRLMTETETEVFNFTQSLRHAMERAFILISNATQNGTVKLEITQNH 299
DB 240 FISRLLEGHYPTDTRLFPENYEIKLSDNGSEFYHAIDRASLLAREGGNNVILKSTGDDVV 299
QY 300 SAHVSNEPVGWNEDLDIVSQSGDLTISNPTVLTIESLKAISKSTVHLSVPRPRTL 359
DB 300 ELSSTSPETGTVKREVDANDVEGSGSLKISFNKYMMDALKADINDEVEFEFGTMKPFIL 359
QY 360 TP-GDESESFQLITPVRT 377
DB 360 KPKGDD-DSVTQLILPRT 376

RESULT 20
ID Q6GD88 PRELIMINARY; PRT; 377 AA.
AC Q6GD88;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE DNA polymerase III, beta chain (EC 2.7.7.7).
GN ORFNames=SA0002;
OS Staphylococcus aureus subsp. aureus MSSA476.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=282459;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=MSSA476;
RA Holden M.T.G., Feil E.J., Lindsay J.A., Peacock S.J., Day N.P.J.,
RA Enright M.C., Foster T.J., Moore C.E., Hurbt L., Atkin R., Barron A.,
RA Bason N., Bentley S.D., Chillingworth C., Chillingworth T.,
RA Churcher C., Clark L., Corton C., Cronin A., Doggett J., Dowd L.,
RA Feltwell T., Hance Z., Harris B., Hauser H., Holroyd S., Jagels K.,
RA James K.D., Lennard N., Line A., Mayes R., Moule S., Mungall K.,
RA Ormond D., Quail M.A., Rabbittowitch E., Rutherford K., Sanders M.,
RA Sharp S., Simmonds M., Stevens K., Whitehead S., Barrett B.G.,
RA Spratt B.G., Parkhill J.;
RC STRAIN=MSSA476;
RA Holden M.T.G., Feil E.J., Lindsay J.A., Peacock S.J., Day N.P.J.,
RA Enright M.C., Foster T.J., Moore C.E., Hurbt L., Atkin R., Barron A.,
RA Bason N., Bentley S.D., Chillingworth C., Chillingworth T.,
RA Churcher C., Clark L., Corton C., Cronin A., Doggett J., Dowd L.,
RA Feltwell T., Hance Z., Harris B., Hauser H., Holroyd S., Jagels K.,
RA James K.D., Lennard N., Line A., Mayes R., Moule S., Mungall K.,
RA Ormond D., Quail M.A., Rabbittowitch E., Rutherford K., Sanders M.,
RA Sharp S., Simmonds M., Stevens K., Whitehead S., Barrett B.G.,
RA Spratt B.G., Parkhill J.;
RT "Complete genomes of two clinical Staphylococcus aureus strains:
RT evidence for the rapid evolution of virulence and drug resistance.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:9786-9791(2004).
DR EMBL; BX571857; CAG41774.1; -

```

```
DR InterPro; IPR001001; DNA_polIII_beta.
DR Pfam; PF00712; DNA_pol3_beta; 1.
DR Pfam; PF02767; DNA_pol3_beta_2; 1.
DR Pfam; PF02768; DNA_pol3_beta_3; 1.
DR SMART; SM00480; POL3BC; 1.
DR TIGRFAMs; TIGR00663; dnan; 1.
KW Nucleotidyltransferase; Transferase.
SQ SEQUENCE 377 AA; 41913 MW; 0A985EF94E044FBC CRC64;

Query Match          39.2%; Score 730.5; DB 2; Length 377;
Best Local Similarity 39.3%; Pred. No. 2.9e-40;
Matches 149; Conservative 90; Mismatches 135; Indels 5; Gaps 4;

QY 1 MIQSFINTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTSGNGQISIENTIPV 60
DB 1 MMEFTIKRDYFITQLNDTLKALSPRTLPILTGKIDAKEHEVILTSGSDSEISIEITPK 60
QY 61 SNEAGLL-ITSPGAILLEASFFINISSLDPDISINVKIEIQHGVLTSGKSEITLKGD 119
DB 61 TVDGEDIVNISSETGSVLPGRFFVDIIKKLPKGVKLTSTNEQFQTLITSGHSEFNLSGLD 120
QY 120 VDQYPRLEQVSTENPLIKTLKLSIIAETAFASLOESRPILTGVHIVLSNHKDFKAVA 179
DB 121 PDQYPLLPQVSRDDAIQLSVKLVKNVIAQINFAVSTSETRPVLTVGNWLIQEN-ELICTA 179
QY 180 TDSHRMSQRLITLNTSADLMVLPKSLRBFSAVFTDDIETVEVFFSPQILFRSEHIS 239
DB 180 TDSHRLAVRKLEQDVSENKNVILPGKALAEINKIMSDNEEDIDIFFASNQVLPKGVN 239
QY 240 FYTRLLEGNYPDTDRLLMTETEVNTQSLRHAMERAPLISNATONGTVKLEITQNH 299
DB 240 FISRLLEGHPDTRLPENYEIKLSIDNGEFYHAIDRASLLAREGGNNVILKSTGDDV 299
QY 300 SAHNSPEVGKVNEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHLSPVRPPTL 359
DB 300 ELSSTSPGIVTKEVDANDVEGSLKISFNKYMMDALKDAINDVEVEFFGTMKPPIL 359
QY 360 TP-GDEESFIQLITPVRT 377
DB 360 KPKGDD--DSVTQLILPIRT 376

RESULT 21
Q6GKU3 PRELIMINARY; PRT; 377 AA.
AC Q6GKU3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE DNA polymerase III, beta chain (EC 2.7.7.7).
GN Name=dnaN; ORFNames=SAR0002;
OS Staphylococcus aureus subsp. aureus MRSA252.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=282459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MRSA252;
RA Holden M.T.G., Feil E.J., Lindsey J.A., Peacock S.J., Day N.P.J.,
RA Enright M.C., Foster T.J., Moore C.E., Hurst L., Atkin R., Barron A.,
RA Bacon N., Bentley S.D., Chillingworth C., Chillingworth T.,
RA Churcher C., Clark L., Cotton C., Cronin A., Doggett J., Dowd L.,
RA Feltwell T., Hance Z., Harris B., Hauser H., Holroyd S., Jagels K.,
RA James K.D., Lennard N., Line A., Mayes R., Moulé S., Mungall K.,
RA Ormond D., Quail M.A., Rabinowitch E., Rutherford K., Sanders M.,
RA Sharp S., Simmonds M., Stevens K., Whitehead S., Bartell B.G.,
RA Spratt B.G., Parkhill J.;
RP "Complete genomes of two clinical Staphylococcus aureus strains:
RT evidence for the rapid evolution of virulence and drug resistance."
RL Proc. Natl. Acad. Sci. U.S.A. 101:9786-9791 (2004).
DR EMBL; BX571856; CAG39030.1; -
DR InterPro; IPR001001; DNA_polIII_beta.
DR Pfam; PF00712; DNA_pol3_beta; 1.
DR Pfam; PF02767; DNA_pol3_beta_2; 1.

DR InterPro; IPR001001; DNA_polIII_beta.
DR Pfam; SM00480; POL3BC; 1.
DR TIGRFAMs; TIGR00663; dnan; 1.
KW Nucleotidyltransferase; Transferase.
SQ SEQUENCE 377 AA; 41913 MW; 0A985EF94E044FBC CRC64;

Query Match          39.2%; Score 730.5; DB 2; Length 377;
Best Local Similarity 39.3%; Pred. No. 2.9e-40;
Matches 149; Conservative 90; Mismatches 135; Indels 5; Gaps 4;

QY 1 MIQSFINTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTSGNGQISIENTIPV 60
DB 1 MMEFTIKRDYFITQLNDTLKALSPRTLPILTGKIDAKEHEVILTSGSDSEISIEITPK 60
QY 61 SNEAGLL-ITSPGAILLEASFFINISSLDPDISINVKIEIQHGVLTSGKSEITLKGD 119
DB 61 TVDGEDIVNISSETGSVLPGRFFVDIIKKLPKGVKLTSTNEQFQTLITSGHSEFNLSGLD 120
QY 120 VDQYPRLEQVSTENPLIKTLKLSIIAETAFASLOESRPILTGVHIVLSNHKDFKAVA 179
DB 121 PDQYPLLPQVSRDDAIQLSVKLVKNVIAQINFAVSTSETRPVLTVGNWLIQEN-ELICTA 179
QY 180 TDSHRMSQRLITLNTSADLMVLPKSLRBFSAVFTDDIETVEVFFSPQILFRSEHIS 239
DB 180 TDSHRLAVRKLEQDVSENKNVILPGKALAEINKIMSDNEEDIDIFFASNQVLPKGVN 239
QY 240 FYTRLLEGNYPDTDRLLMTETEVNTQSLRHAMERAPLISNATONGTVKLEITQNH 299
DB 240 FISRLLEGHPDTRLPENYEIKLSIDNGEFYHAIDRASLLAREGGNNVILKSTGDDV 299
QY 300 SAHNSPEVGKVNEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHLSPVRPPTL 359
DB 300 ELSSTSPGIVTKEVDANDVEGSLKISFNKYMMDALKDAINDVEVEFFGTMKPPIL 359
QY 360 TP-GDEESFIQLITPVRT 377
DB 360 KPKGDD--DSVTQLILPIRT 376

RESULT 22
Q8EU87 PRELIMINARY; PRT; 378 AA.
AC Q8EU87;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE DNA-directed DNA polymerase III beta chain (EC 2.7.7.7).
GN Name=dnaN; OrderedLocusNames=OH00002;
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831;
RX MEDLINE=2220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments."
RL Nucleic Acids Res. 30:3927-3935 (2002).
DR EMBL; AP004593; BAC11958.1; -
DR GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003897; F:DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006260; F:DNA replication; IEA.
DR InterPro; IPR001001; DNA_polIII_beta.
DR Pfam; PF00712; DNA_pol3_beta; 1.
DR Pfam; PF02767; DNA_pol3_beta_2; 1.
DR Pfam; PF02768; DNA_pol3_beta_3; 1.
DR SMART; SM00480; POL3BC; 1.
DR TIGRFAMs; TIGR00663; dnan; 1.
KW Complete proteome; DNA-directed DNA polymerase;
```



```

KW Nucleotidyltransferase; Transferase.
SQ SEQUENCE 378 AA; 42231 MW; BABACD3AE3BD962 CRC64;

Query Match 38.8%; Score 722.5; DB 2; Length 378;
Best Local Similarity 41.1%; Pred. No. 9.8e-40;
Matches 157; Conservative 85; Mismatches 129; Indels 11; Gaps 7;

QY 2 IQFSINRILFHALNTTKRAISTKNAIPILSSIKIEVTSSTGTLTSGNGQISIENTIPVS 61
Db 1 MRFTIQDKLINGVNMKAISARTVPIILTMKIEVKNHGVTLTGSDSDISIEYIPI- 59
QY 62 NENAGLLIT---SPCAILLEAFNIISSLPDISINVKIEHQHVLTSGKSEITLKGK 118
Db 60 -EEDGIVHENIEETIILQAKYFFDIWKLPSTVDIVDDQLMVRITSGKAENLNGQ 118
QY 119 DVDQYPRLOVSTENPLIKTKLLKSIIAETAFASLOESRPILTVGHVILSNHKDFRAV 178
Db 119 SAEYEPQLPKVQNTSEFELPIDLLAKSMIKQTVFAVSTMETRPILTVGNLKLVDN-SLSFT 177
QY 179 ATDSHRMSORLITLNTSADL-MVVLPSKSLREFSAVFTDDIETVEVFFSPSOLFRSEH 237
Db 178 ATDSHRLARREITPVSNAPIEISQIVVPGKSLMELNKILGDSSEETVEISVTNNQILFRTH 237
QY 238 ISFYTRLLEGNYPDTRLMTTEFETEVFNQSLRHAMERAFILSNATQNGTVKLEITQN 297
Db 238 LNFLRLGNGNPETSRILPEOSKTKIQIKTKELGTIDRASLLAKEERNNVKFNAPGN 297
QY 298 H-ISAHVNSPEVGVKNEIDLIVSQSGDLTTSFNPTYLIESKAIKSETVKIHFLSPVPR 356
Db 298 SMIEISSNSPEVGVNVEITADQMEGDVKISFSKYMIDALKALEYDEVQIEFTGANRP 357
QY 357 FTLTP-GDEESFIQLITPVRT 377
Db 358 FIIRPVGD--DSILQLITPVRT 377

RESULT 23
Q92FV1 PRELIMINARY; PRT; 381 AA.
AC Q92FV1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE DNA polymerase III, beta chain.
GN Name=dnaN; OrderedLocusNames=lin0002;
OC Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OC NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krest J., Kuhn M., Kunst F., Kurapkat G.,
RA Madieno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluerer T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596163; CAC95235.1; -.
DR PIR; AC1433; AC1433.
DR ListList; LIN0002; -.
DR GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR InterPro; IPR001001; DNA polIII_beta.

DR Pfam; PF00712; DNA_pol3_beta; 1.
DR Pfam; PF02767; DNA_pol3_beta_2; 1.
DR Pfam; PF02768; DNA_pol3_beta_3; 1.
DR SMART; SM00480; POL3BC; 1.
DR TIGRFAMs; TIGR00663; dnan; 1.
KW Complete proteome.
SQ SEQUENCE 381 AA; 42497 MW; 977614257898F46D CRC64;

Query Match 38.1%; Score 710; DB 2; Length 381;
Best Local Similarity 39.4%; Pred. No. 6.7e-39;
Matches 150; Conservative 92; Mismatches 133; Indels 6; Gaps 5;

QY 2 IQFSINRILFHALNTTKRAISTKNAIPILSSIKIEVTSSTGTLTSGNGQISIENTIP-V 60
Db 1 MKFVIERDLVQAVNEVTRAIASRTTIPILTGKIVNVDEGVTLTGSDSDISIEAFIPLI 60
QY 61 SNEAGLLITSPGAILLEASFFINIISLPDISINVKIEHQHVLTSGKSEITLKGKDV 120
Db 61 ENDEVIVEVESFGGIVLQSKYFGDIVRRLPRENVIEVTTNYQTNISSGQASFTLGLDP 120
QY 121 DOYPRLOVSTENPLIKTKLLKSIIAETAFASLOESRPILTVGHVILSNHKDFRAV 180
Db 121 MEYKPLPEVTDGKNIKIPINVLKNIIRQTVFAVSAIEVRPVLTVGNWIIKDNK-LSAVAT 179
QY 181 DSHRMSORLITLNTSADL-MVVLPSKSLREFSAVFTDDIETVEVFFSPSOLFRSEHIS 239
Db 180 DSHRLARREITPVSNAPIEISQIVVPGKSLMELNKILGDSSEETVEISVTNNQILFRTH 239
QY 240 FYTRLLEGNYPDTRLMTTEFETEVFNQSLRHAMERAFILSNATQNGTVKLEITQN-H 298
Db 240 FYSLRLGEGYDTSRLIPTDKSELVINSRAFLQIDRASLLAKEERNNVKFNAPGN 299
QY 299 ISAHVNSPEVGVKNEIDLIVSQSGDLTTSFNPTYLIESKAIKSETVKIHFLSPVPRPT 358
Db 300 VEVSNSPEVGVNVEITADQMEGDVKISFSKYMIDALKALEYDEVQIEFTGANRP 359
QY 359 LTPGD--EESFIQLITPVRT 377
Db 360 LRPKDATNPNEILQLITPVRT 380

RESULT 24
Q8YAW1 PRELIMINARY; PRT; 381 AA.
AC Q8YAW1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE DNA polymerase III, beta chain.
GN Name=dnaN; OrderedLocusNames=lmo0002;
OC Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OC NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krest J., Kuhn M., Kunst F., Kurapkat G.,
RA Madieno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluerer T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591973; CAC98217.1; -.
DR PIR; AC1432; AC1432.
DR ListList; LMO0002; -.

```



GO: 0008408; F:3'-5' exonuclease activity; IEA.  
 DR GO: 0003677; F:DNA binding; IEA.  
 DR GO: 0003887; F:DNA-directed DNA polymerase activity; IEA.  
 DR GO: 0006260; P:DNA replication; IEA.  
 DR InterPro: IPR001001; DNA\_polIII\_beta.  
 DR Pfam: PF00712; DNA\_pol3\_beta\_1.  
 DR Pfam: PF02767; DNA\_pol3\_beta\_2.  
 DR Pfam: PF02768; DNA\_pol3\_beta\_3.  
 DR SMART: SM00480; POL3BC\_1.  
 DR TIGRFAMs: TIGR00663; dnan.  
 KW Complete proteome.  
 SQ SEQUENCE 381 AA; 42429 MW; 49F34BBEDBFB7844 CRC64;

Query Match 37.7%; Score 703; DB 2; Length 381;  
 Best Local Similarity 38.8%; Pred. No. 1.9e-38;  
 Matches 148; Conservative 93; Mismatches 134; Indels 6; Gaps 5;

QY 2 IQSINRTLFHAIHNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNISIENTIP-V 60  
 DB 1 MKFVIERDLVQAVNEVTRAIARTTIPILTKIKIVNDEGVTLTGSDSDISIEAFIPI 60  
 QY 61 SNENAGLLITSPGAILLEASFFNIISLPDISINKEIEHQHVLTSGKSEITLKGDV 120  
 DB 61 ENDEVIVESEFGGIVLQSKYFGDIVRRLPEENVEIVTSNYQTNISSQASFTLGLDP 120  
 QY 121 DQYPRLOEVSTENPLILTKLLKSIIAETAFASLQESRPILTGVHIVLSNHNKDFKAVAT 180  
 DB 121 MEYPKLPEVTDGKTIKIPINVLKNIVRQTVFAVSAIEVRPVLTVGNWIIKENK-LSAVAT 179  
 QY 181 DSHRMSQRLITLD-NTSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHS 239  
 DB 180 DSHRLALREIPLETIDIEEYNIPIPKSLSELNKLDDASESIEMTLANNQILFKLDLL 239  
 QY 240 FYTRLLEGNYPTDRLMTPEFEVVTQSLRHAMERAFILSNATONGTVKLEITON-H 298  
 DB 240 FYSRLLEGSYPTSLIPDTKSELVINSKAFLOAIDRSLARENRVNVIKMTLENGQ 299  
 QY 299 ISAHVNSPEVGKVNEDLDIVSOGSDLTISFNPTYLIESLKAISKSETVKIHLSPVRPFT 358  
 DB 300 VEVSSNSPEVGNVSENVFSQSFTGEEKISFNKYMMDALRAFEQDDIQISFGTMRPFV 359  
 QY 359 LTPGD--EESFQILTPVPT 377  
 DB 360 LRPKDAANPNEILQLITPVRT 380

RESULT 25  
 Q725G9  
 ID Q725G9 PRELIMINARY; PRT; 381 AA.  
 AC Q725G9;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE DNA polymerase III, beta subunit (EC 2.7.7.7).  
 GN Name=dnan; OrderedLocNames=LMOF2365\_0002;  
 OS Listeria monocytogenes (serotype 4b, strain F2365).  
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
 OX NCBI\_TaxID=265669;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC PubMed=15115801; DOI=10.1093/nar/gkh562;  
 RA Nelson K.E., Fouts D.E., Mongodin E.F., Ravel J., DeBoy R.T.,  
 RA Kolonay J.F., Rasko D.A., Angiuoli S.V., Gill S.R., Paulsen I.T.,  
 RA Peterson J.D., White O., Nelson W.C., Nierman W.C., Beanan M.J.,  
 RA Brinkac L.M., Daugherty S.C., Dodson R.J., Durkin A.S., Madupu R.,  
 RA Haft D.H., Selengut J., Van Aken S.E., Khouri H.M., Fedorova N.,  
 RA Forberger H.A., Tran B., Katharou S., Wonderling L.D., Uhlrich G.A.,  
 RA Bayles D.O., Luchansky J.B., Fraser C.M.;  
 RT "Whole genome comparisons of serotype 4b and 1/2a strains of the food-  
 borne pathogen *Listeria monocytogenes* reveal new insights into the  
 core genome components of this species."  
 RL Nucleic Acids Res. 32:2386-2395(2004).  
 DR EMBL; AE017322; AAT02792.1; -.

DR GO: 0003887; F:DNA-directed DNA polymerase activity; IEA.  
 DR GO: 0016740; F:transferase activity; IEA.  
 DR InterPro: IPR001001; DNA\_polIII\_beta.  
 DR Pfam: PF00712; DNA\_pol3\_beta\_1.  
 DR Pfam: PF02767; DNA\_pol3\_beta\_2.  
 DR Pfam: PF02768; DNA\_pol3\_beta\_3.  
 DR SMART: SM00480; POL3BC\_1.  
 DR TIGRFAMs: TIGR00663; dnan.  
 KW Complete proteome; Nucleotidyltransferase; Transferase.  
 SQ SEQUENCE 381 AA; 42429 MW; 49F34BBEDBFB7844 CRC64;

Query Match 37.7%; Score 703; DB 2; Length 381;  
 Best Local Similarity 38.8%; Pred. No. 1.9e-38;  
 Matches 148; Conservative 93; Mismatches 134; Indels 6; Gaps 5;

QY 2 IQSINRTLFHAIHNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNISIENTIP-V 60  
 DB 1 MKFVIERDLVQAVNEVTRAIARTTIPILTKIKIVNDEGVTLTGSDSDISIEAFIPI 60  
 QY 61 SNENAGLLITSPGAILLEASFFNIISLPDISINKEIEHQHVLTSGKSEITLKGDV 120  
 DB 61 ENDEVIVESEFGGIVLQSKYFGDIVRRLPEENVEIVTSNYQTNISSQASFTLGLDP 120  
 QY 121 DQYPRLOEVSTENPLILTKLLKSIIAETAFASLQESRPILTGVHIVLSNHNKDFKAVAT 180  
 DB 121 MEYPKLPEVTDGKTIKIPINVLKNIVRQTVFAVSAIEVRPVLTVGNWIIKENK-LSAVAT 179  
 QY 181 DSHRMSQRLITLD-NTSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHS 239  
 DB 180 DSHRLALREIPLETIDIEEYNIPIPKSLSELNKLDDASESIEMTLANNQILFKLDLL 239  
 QY 240 FYTRLLEGNYPTDRLMTPEFEVVTQSLRHAMERAFILSNATONGTVKLEITON-H 298  
 DB 240 FYSRLLEGSYPTSLIPDTKSELVINSKAFLOAIDRSLARENRVNVIKMTLENGQ 299  
 QY 299 ISAHVNSPEVGKVNEDLDIVSOGSDLTISFNPTYLIESLKAISKSETVKIHLSPVRPFT 358  
 DB 300 VEVSSNSPEVGNVSENVFSQSFTGEEKISFNKYMMDALRAFEQDDIQISFGTMRPFV 359  
 QY 359 LTPGD--EESFQILTPVPT 377  
 DB 360 LRPKDAANPNEILQLITPVRT 380

RESULT 26  
 AAT02792  
 ID AAT02792 PRELIMINARY; PRT; 381 AA.  
 AC AAT02792;  
 DT 10-MAY-2004 (TrEMBLrel. 27, Created)  
 DT 10-MAY-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 10-MAY-2004 (TrEMBLrel. 27, Last annotation update)  
 DE DNA polymerase III, beta subunit (EC 2.7.7.7).  
 GN Name=OR LMOF2365\_0002;  
 OS Listeria monocytogenes str. 4b F2365.  
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria;  
 OX NCBI\_TaxID=265669;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=4b F2365;  
 RX PubMed=15115801;  
 RA Nelson K.E., Fouts D.E., Mongodin E.F., Ravel J., DeBoy R.T.,  
 RA Kolonay J.F., Rasko D.A., Angiuoli S.V., Gill S.R., Paulsen I.T.,  
 RA Peterson J.D., White O., Nelson W.C., Nierman W.C., Beanan M.J.,  
 RA Brinkac L.M., Daugherty S.C., Dodson R.J., Durkin A.S., Madupu R.,  
 RA Haft D.H., Selengut J., Van Aken S., Khouri H., Fedorova N.,  
 RA Forberger H., Tran B., Katharou S., Wonderling L.D., Uhlrich G.A.,  
 RA Bayles D.O., Luchansky J.B., Fraser C.M.;  
 RT "Whole genome comparisons of serotype 4b and 1/2a strains of the food-  
 borne pathogen *Listeria monocytogenes* reveal new insights into the  
 core genome components of this species."  
 RL Nucleic Acids Res. 32:2386-2395(2004).  
 DR EMBL; AE017322; AAT02792.1; -.

Qy	62	NEAGLL-ITSPGAILLEASFFINIISLPDISINVKEIQHVV-LTSGKSEITLKGKD	119
Db	61	EDGKEIVEVKQSGSIVLOAKYFSEIVKLPKETVEI-SVENHLMTKITSGKSEFNGLJD	119
Qy	120	VDQYPRLOEVSTENPLIKTKLLKSIITAETAFAASLOESRPLITGVHIVLSNHNKDFKAVA	179
Db	120	SASYPLLPQLEEBHHVFKIPTDLLKHMIRQTVEFVAVSTSETRPILTGNNWKNVY-SEUTCIA	178
Qy	180	TDHRMSORLITIDNTSADLM--VVLPSKLSUREFSAVFTDDIETVEVFPSPQILFRSEH	237
Db	179	TDSHRLAIRKAKTEGIVDEQANVVIKGSLSNELSKILDESEWVDIVITEYQVLFRTKH	238
Qy	238	ISFYTRLEGGYPTDRLMLTEFTEVVFVNTQSLRHAMERAPLISNATQNGTVKLE-ITQ	296
Db	239	LLFFSRLEGGYPTDTRLIPAESKTDIFVNTKEFLQADIRASLLARDGRNNVVKLSLTLR	298
Qy	297	NHISAHVNSPEVGVKNEDLDIVSOGSDLTISFNPTLYLESLSKAIKSETVKHIFLSPVRP	356
Db	299	AMLEISSNSPEIGKVBEVOCEKVDGELKISFSAKYNMDALKDSTEIKISFTGAMRP	358
Qy	357	FTLTPGDEESFIQILTPVRT	377
Db	359	FLIRTVN-DESIQLILPVRT	378
RESULT	28		
Q73FK4		PRELIMINARY;	PRT; 379 AA.
AC	Q73FK4		
DT	05-JUL-2004	(TrEMBLrel. 27, Created)	
DT	05-JUL-2004	(TrEMBLrel. 27, Last sequence update)	
DT	05-JUL-2004	(TrEMBLrel. 27, Last annotation update)	
DE	DNA polymerase III, beta subunit (EC 2.7.7.7).		
GN	Name=dnan, orderedlocusNames=BCE0002;		
OS	Bacillus cereus (strain ATCC 10987).		
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.		
OX	NCBI_TaxID=222523;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	PubMed=14960714;		
RA	Rasko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,		
RA	Shores K.A., Fouts D.E., Tourasse N.J., Angluoli S.V., Kolonay J.F.,		
RA	Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;		
RT	"The genome sequence of <i>Bacillus cereus</i> ATCC 10987 reveals metabolic		
RT	adaptations and a large plasmid related to <i>Bacillus anthracis</i> pX01."		
RL	Nucleic Acids Res. 32:977-988(2004).		
DR	EMBL; AE017264; AAS38938.1; -		
DR	TIGR; BC80002;		
DR	GO; GO:0003987; F:DNA-directed DNA polymerase activity; IEA.		
DR	GO; GO:0016740; F:transferase activity; IEA.		
DR	InterPro; IPR001001; DNA_poliIII_beta.		
DR	Pfam; PF00712; DNA_poli3_beta_1.		
DR	Pfam; PF02767; DNA_poli3_beta_2; 1.		
DR	Pfam; PF02768; DNA_poli3_beta_3; 1.		
DR	SMART; SM00480; POL3Bc; 1.		
DR	TIGRFAMS; TIGR00663; dnan; 1.		
SK	Complete proteome; Nucleotidyltransferase; Transferase.		
QW	SEQUENCE 379 AA; 42529 MW; A413B2CFD197C439 CRC64;		
Query Match	37.2%;	Score 693;	DB 2; Length 379;
Best Local Similarity	38.6%;	Pred. No. 8.9e-38;	
Matches	147;	Conservative	86; Mismatches 140; Indels 8; Gaps 7
Qy	2	IQFSINRTLFHAIHTTKRAISTKNAIPILSIIKIEVTSTGVTLTGSGNQGISIENTIPVS	61
Db	1	MRFTIQDYLVRSQVDYMKAVSSRTTIPILTGKIVKVAITEEGVTLTGSDADISIESFIPVE	60
Qy	62	NEAGLL-ITSPGAILLEASFFINIISLPDISINVKEIQHVV-LTSGKSEITLKGKD	119
Db	61	ENGKEIVEVKQSGSIVLOAKYFSEIVKLPKETVEI-SVENHLMTKITSGKSEFNGLJD	119
Qy	120	VDQYPRLOEVSTENPLIKTKLLKSIITAETAFAASLOESRPLITGVHIVLSNHNKDFKAVA	179

Db 120 SAEPYLLPQIEEHVFKIPTDLKHMIRQVFAVSTSETRPILTVGNWKYN-SELTGIA 178  
 QY 180 TDSHRMSORLITLNTSADLM--VVLPSKSLRFSVFTDDIETVEVFFSPQILFRSEH 237  
 Db 179 TDSHRLALRKAKIEGIADEFOANVVIPEGKSLNELSKILDESEEMVDIVITEYQVLFRTKH 238  
 QY 238 ISFYTRLLEGNYPTDRLMTPEFETEVVNTQSLRHAMERAFILSNATONGTVKLE-ITQ 296  
 Db 239 LFFSRLLEGNYPTDRLIPAESKTDIFVNTKEFLQAIADRASLARDGRNNVVKLSTLEQ 298  
 QY 297 NHISAHVNSPEVQKVEDLDIVSQSGDLTISFNPTYLIESLKAISKSETVKIHLFSPVRP 356  
 Db 299 AMLEISSNSPEIGKVVEEQCEKVDGEELKISFSKAKYMDALKALDSTEIKISFTGMARP 358  
 QY 357 FTLTGDEESFQILTPVRT 377  
 Db 359 FLIRTVN-DESIQILTPVRT 378

## RESULT 29

AAS38938  
 ID AAS38938 PRELIMINARY; PRT; 379 AA.  
 AC AAS38938  
 DT 02-MAR-2004 (TREMELrel. 27, Created)  
 DT 02-MAR-2004 (TREMELrel. 27, Last sequence update)  
 DE DNA polymerase III, beta subunit (EC 2.7.7.7).  
 GN DNAN OR BC00002.  
 OS *Bacillus cereus* (strain ATCC 10987).  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Bacillus*.  
 OX NCBI\_TaxID=222523;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=14960714;  
 RA Rasko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L., Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F., Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;  
 RT "The genome sequence of *Bacillus cereus* ATCC 10987 reveals metabolic adaptations and a large plasmid related to *Bacillus anthracis* pXOI.";  
 RL Nucleic Acids Res. 32:977-988(2004).  
 DR EMBL; AE017264; AAS38938.1; -;  
 DR TIGR; BC00002; -;  
 KW Nucleotidyltransferase; Transferase.  
 SQ SEQUENCE 379 AA; 42529 MW; A413B2CFD197C439 CRC64;

Query Match 37.2%; Score 693; DB 2; Length 379;  
 Best Local Similarity 38.6%; Pred. No. 8, 9e-38;  
 Matches 147; Conservative 86; Mismatches 140; Indels 8; Gaps 7;  
 QY 2 IQFSINRILFIHALNTTKRAISTKNAIPILSIKIEVTSTGTTLTGSNGQISIENTIPVS 61  
 Db 1 MRFTIQDYLVRSDQVDMKAVSSRTTIPILGKIVATEEGVTLTGSADISIESFIPVE 60  
 QY 62 NENAGLL-ITSPGALLLRASFINIISLDPISINKEIOHQV-ITSGKSEITLKGKD 119  
 Db 61 ENGEIVSVKSGSGLVLOAKYFSEIVKLPKETVEI-SVENHLMTKITSGKSEFNGLD 119  
 QY 120 VDQYPRLOEVSTENPLILKTKLSIIAETAFAASLQESRPILTVGHVILSNHDKPKAVA 179  
 Db 120 SAEPYLLPQIEEHVFKIPTDLKHMIRQVFAVSTSETRPILTVGNWKYN-SELTGIA 178  
 QY 180 TDSHRMSORLITLNTSADLM--VVLPSKSLRFSVFTDDIETVEVFFSPQILFRSEH 237  
 Db 179 TDSHRLALRKAKIEGIADEFOANVVIPEGKSLNELSKILDESEEMVDIVITEYQVLFRTKH 238  
 QY 238 ISFYTRLLEGNYPTDRLMTPEFETEVVNTQSLRHAMERAFILSNATONGTVKLE-ITQ 296  
 Db 239 LFFSRLLEGNYPTDRLIPAESKTDIFVNTKEFLQAIADRASLARDGRNNVVKLSTLEQ 298  
 QY 297 NHISAHVNSPEVQKVEDLDIVSQSGDLTISFNPTYLIESLKAISKSETVKIHLFSPVRP 356  
 Db 299 AMLEISSNSPEIGKVVEEQCEKVDGEELKISFSKAKYMDALKALDSTEIKISFTGMARP 358

QY 357 FTLTGDEESFQILTPVRT 377  
 Db 359 FLIRTVN-DESIQILTPVRT 378  
 RESULT 30  
 Q81W34  
 ID Q81W34 PRELIMINARY; PRT; 379 AA.  
 AC Q81W34; Q61537; Q6KYT0;  
 DT 01-JUN-2003 (TREMELrel. 24, Created)  
 DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)  
 DT 01-OCT-2004 (TREMELrel. 28, Last annotation update)  
 DE DNA polymerase III, beta subunit.  
 GN Name=dnan-1; OrderedLocusNames=BA0002, BAS0002; ORFNames=GBAA0002;  
 OS *Bacillus anthracis*.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Bacillus*.  
 OX NCBI\_TaxID=1392;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ames / isolate Porton;  
 RX MEDLINE=22608414; PubMed=12721629; DOI=10.1038/nature01586;  
 RA Read T.D., Peterson S.N., Tourasse N.J., Baillie L.W., Paulsen I.T., Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R., Holtzapple E.K., Okstad O.A., Helgason E., Ristone J., Wu M., Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.L., DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H., Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D., Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F., Berty K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C., Hazen A., Cline R.T., Redmond C., Thwaite J.E., White O., Salzberg S.L., Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolstoe A.-B., Fraser C.M.;  
 RA "The genome sequence of *Bacillus anthracis* and comparison to closely related bacteria.";  
 RT Nature 423:81-86(2003).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ames / isolate 0581;  
 RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B., Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L., Fraser C.M.;  
 RT "Bacillus anthracis comparative genomics";  
 RN Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.  
 RP [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Stearne;  
 RA Brettin T.S., Bruce D., Chailacombe J.F., Gilna P., Han C., Hill K., Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R., Richardson P., Rubin E., Tice H.;  
 RA Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE017024; AAP24060.1; -;  
 DR EMBL; AE017334; AAP24060.1; -;  
 DR EMBL; AB017225; AAT52344.1; -;  
 DR TIGR; BA0002; -;  
 DR GO; GO:0008408; F.3'-5' exonuclease activity; IEA.  
 DR GO; GO:0003677; F.DNA binding; IEA.  
 DR GO; GO:0003887; F.DNA-directed DNA polymerase activity; IEA.  
 DR GO; GO:0008260; P.DNA replication; IEA.  
 DR InterPro; IPR001001; DNA polIII beta.  
 DR Pfam; PF00712; DNA pol3\_beta; 1.  
 DR Pfam; PF02767; DNA pol3\_beta\_2; 1.  
 DR Pfam; PF02768; DNA pol3\_beta\_3; 1.  
 DR SMART; SM00480; POL3Bc; 1.  
 DR TIGRFAMS; TIGR00663; dnan; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 379 AA; 42650 MW; C49D709DDBA03452E CRC64;

Query Match 37.1%; Score 692; DB 2; Length 379;  
 Best Local Similarity 38.8%; Pred. No. 1e-37;  
 Matches 148; Conservative 84; Mismatches 141; Indels 8; Gaps 7;  
 QY 2 IQFSINRILFIHALNTTKRAISTKNAIPILSIKIEVTSTGTTLTGSNGQISIENTIPVS 61

```
Db 1 MRFSQKDYLRVSQDVQVKAQVSRFTTIPILTGKIKVVAETEGVTLLGSDADISIESFIPVE 60
Qy 62 NENAGLL-ITSPGAILLEASFPINISSLPDISINVKIEIOHQVV-LTSGKSEITLKGD 119
Db 61 EDGKEIVEVKQSGSVLQAKYFSEIVKLPKRETVET-SVENHLMTKITSGKSEFNGLD 119
Qy 120 VDQYPRLOEVSTENPLIKTKLLKIIAETAPASLQSRPILITGVHVLGNHDKFKAVA 179
Db 120 SAEYPLLPQIEHHVFKIPTDLLKHMIRQTVFAVSTSETRPILITGVNKKVYN-SELTGIA 178
Qy 180 TDSHRMSORLITLDNTSADLM--VVLPSKSLREFSAVFTDDIETVEVFFSPQILFRSEH 237
Db 179 TDSHRLAURKAKIEGIVDFQANVVPKGSNLSKILDESEMVDIVITEVQVLFRTKH 238
Qy 238 ISFYTRLLEGNYPDTRILLMTTEFEVFNVTQSLRHAMERAPLISNATQNGTVKLE-ITQ 296
Db 239 LLFPSRLLEGNYPDTRILIPAESKTDIFVNTKEFLQADRSLARDGRNNVVKLSTLEQ 298
Qy 297 NHISAHVNSPEVKKNEDLDIVSQSGDLTISFNPTYLIESLKAIKSTVKIHFSPVRP 356
Db 299 AMLEISSNSPEIGKVEEQCEKVDGELKISFSKAYMMMDALKALDSTEIKISFTGMRP 358
Qy 357 FTLTPGDEESFIQILTPVRT 377
Db 359 FLIRTVN-DESIILQILPVRT 378

RESULT 31
AAT29080
ID AAT29080 PRELIMINARY; PRT; 379 AA.
AC AAT29080;
DT 01-JUN-2004 (TReMBLrel. 27, Created)
DT 01-JUN-2004 (TReMBLrel. 27, Last sequence update)
DT 01-JUN-2004 (TReMBLrel. 27, Last annotation update)
DE DNA polymerase III, beta subunit.
GN DNAN-1 OR GBAA0002.
OS Bacillus anthracis str. Ames 0581.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus cereus group; Bacillus anthracis.
OX NCBI_TaxID=261594;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames 0581;
RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S., Fraser C.M.;
RT "Bacillus anthracis comparative genomics.";
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017334; AAT29080.1; -- C49D709DDBA03452E CRC64;
SQ SEQUENCE 379 AA; 42650 MW; 42650 MW; 42650 MW;

Query Match 37.1%; Score 692; DB 2; Length 379;
Best Local Similarity 38.8%; Pred. No. 1e-37;
Matches 148; Conservative 84; Mismatches 141; Indels 8; Gaps 7;

Qy 2 IOFSINRFLFIHALMTTKRAISTKNAIPILSIIKIEVSTGTVTLGSGNQISIENTIPVS 61
Db 1 MRFSQKDYLRVSQDVQVKAQVSRFTTIPILTGKIKVVAETEGVTLLGSDADISIESFIPVE 60
Qy 62 NENAGLL-ITSPGAILLEASFPINISSLPDISINVKIEIOHQVV-LTSGKSEITLKGD 119
Db 61 EDGKEIVEVKQSGSVLQAKYFSEIVKLPKRETVET-SVENHLMTKITSGKSEFNGLD 119
Qy 120 VDQYPRLOEVSTENPLIKTKLLKIIAETAPASLQSRPILITGVHVLGNHDKFKAVA 179
Db 120 SAEYPLLPQIEHHVFKIPTDLLKHMIRQTVFAVSTSETRPILITGVNKKVYN-SELTGIA 178
Qy 180 TDSHRMSORLITLDNTSADLM--VVLPSKSLREFSAVFTDDIETVEVFFSPQILFRSEH 237
Db 179 TDSHRLAURKAKIEGIVDFQANVVPKGSNLSKILDESEMVDIVITEVQVLFRTKH 238
Qy 238 ISFYTRLLEGNYPDTRILLMTTEFEVFNVTQSLRHAMERAPLISNATQNGTVKLE-ITQ 296
Db 239 LLFPSRLLEGNYPDTRILIPAESKTDIFVNTKEFLQADRSLARDGRNNVVKLSTLEQ 298
Qy 297 NHISAHVNSPEVKKNEDLDIVSQSGDLTISFNPTYLIESLKAIKSTVKIHFSPVRP 356
Db 299 AMLEISSNSPEIGKVEEQCEKVDGELKISFSKAYMMMDALKALDSTEIKISFTGMRP 358
Qy 357 FTLTPGDEESFIQILTPVRT 377
Db 359 FLIRTVN-DESIILQILPVRT 378

RESULT 32
DP3B_BACSU
ID DP3B_BACSU STANDARD; PRT; 378 AA.
AC P05649; P11571;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE DNA polymerase III, beta chain (EC 2.7.7.7).
GN Name=dnaN; Synonyms=dnaG; Ordered locus names=BSU00020;
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=85215612; PubMed=2987847;
RA Moriya S., Ogasawara N., Yoshikawa H.;
RT "Structure and function of the region of the replication origin of the
RT Bacillus subtilis chromosome. III. Nucleotide sequence of some 10,000
RT base pairs in the origin region.";
RL Nucleic Acids Res. 13:2251-2265(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RT subtilis chromosome containing the replication origin.";
RL DNA Res. 1:1-14(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RC MEDLINE=98044033; PubMed=9384377; DOI=10.1038/36786;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Enrian K.-D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M.,
RA Jones L.-M., Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M.,
RA Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S.,
RA Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J.,
RA Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C.,
RA Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S.,
RA Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B.,
RA Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S.,
RA Prescott A.M., Pressecan E., Pujic P., Purnelle B., Rapoport G.,
RA Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B.,
RA Rose M., Sadaie Y., Sato T., Scanlan E., Schleich S., Schroeter R.,
RA Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serror P.,
RA Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H.,
RA Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P.,
RA Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F.,
RA Vassartotti A., Viari A., Wambutt R., Wedler E., Wedler H.,
RA Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K.,
RA Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E.,
RA Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
```

```
Db 239 LLFPSRLLEGNYPDTRILIPAESKTDIFVNTKEFLQADRSLARDGRNNVVKLSTLEQ 298
Qy 297 NHISAHVNSPEVKKNEDLDIVSQSGDLTISFNPTYLIESLKAIKSTVKIHFSPVRP 356
Db 299 AMLEISSNSPEIGKVEEQCEKVDGELKISFSKAYMMMDALKALDSTEIKISFTGMRP 358
Qy 357 FTLTPGDEESFIQILTPVRT 377
Db 359 FLIRTVN-DESIILQILPVRT 378
```

```
RESULT 32
DP3B_BACSU
ID DP3B_BACSU STANDARD; PRT; 378 AA.
AC P05649; P11571;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 01-OCT-2004 (Rel. 45, Last annotation update)
DE DNA polymerase III, beta chain (EC 2.7.7.7).
GN Name=dnaN; Synonyms=dnaG; Ordered locus names=BSU00020;
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=85215612; PubMed=2987847;
RA Moriya S., Ogasawara N., Yoshikawa H.;
RT "Structure and function of the region of the replication origin of the
RT Bacillus subtilis chromosome. III. Nucleotide sequence of some 10,000
RT base pairs in the origin region.";
RL Nucleic Acids Res. 13:2251-2265(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RT subtilis chromosome containing the replication origin.";
RL DNA Res. 1:1-14(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RC MEDLINE=98044033; PubMed=9384377; DOI=10.1038/36786;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Enrian K.-D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M.,
RA Jones L.-M., Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M.,
RA Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S.,
RA Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J.,
RA Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C.,
RA Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S.,
RA Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B.,
RA Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S.,
RA Prescott A.M., Pressecan E., Pujic P., Purnelle B., Rapoport G.,
RA Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B.,
RA Rose M., Sadaie Y., Sato T., Scanlan E., Schleich S., Schroeter R.,
RA Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serror P.,
RA Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H.,
RA Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P.,
RA Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F.,
RA Vassartotti A., Viari A., Wambutt R., Wedler E., Wedler H.,
RA Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K.,
RA Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E.,
RA Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
```

```

RT subtilis.";
RL Nature 390:249-256(1997).
RN [4]
RP SEQUENCE OF 1-29 FROM N.A.
RC STRAIN=168 / PSL1;
RX MEDLINE=89030659; PubMed=2846289;
RA "Moriya S., Fukuoka T., Ogasawara N., Yoshikawa H.;
RT Regulation of Initiation of the Chromosomal Replication by DnaA-boxes
in the origin region of the Bacillus subtilis chromosome.";
RL EMBO J. 7:2911-2917(1988).
CC CC
CC -!- FUNCTION: DNA polymerase III is a complex, multichain enzyme
responsible for most of the replicative synthesis in bacteria.
CC This DNA polymerase also exhibits 3' to 5' exonuclease activity.
CC The beta chain is required for initiation of replication once it
CC is clamped onto DNA, it slides freely (bidirectional and ATP-
CC independent) along duplex DNA.
CC CC
CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
+ {DNA} (N).
CC CC
CC -!- SUBUNIT: Contains a core (composed of alpha, epsilon, and theta
chains) that can repair short gaps created by nuclease in duplex
CC DNA. For efficient replication of the long, single-stranded
CC templates, pol III requires the auxiliary chains beta, gamma, and
CC delta.
CC CC
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X02369; CAAG26218.1; -
DR EMBL; D26185; BAA05238.1; -
DR EMBL; Z99104; CAB11778.1; -
DR EMBL; X12779; CAA31271.1; -
DR PUR; B22930; B22930.
DR Subtilisin; EG10066; dnaN.
DR InterPro; IPR001001; DNA_polIII_beta.
DR Pfam; PF00712; DNA_pol3_beta_1.
DR DR PFam; PF02767; DNA_pol3_beta_2; 1.
DR DR PFam; PF02768; DNA_pol3_beta_3; 1.
DR DR SMART; SM00480; POL3BG; 1.
DR DR TIGRfamS; TIGR00663; dnaN; 1.
KW Complete proteome; DNA replication; DNA-directed DNA polymerase;
KW Transference.
SQ SEQUENCE 378 AA; 42103 MW; C1F040B5D53026C8 CRC64;

Query Match 36.5%; Score 680.5; DB 1; Length 378;
Best Local Similarity 37.1%; Pred. No. 6e-37;
Matches 141; Conservative 90; Mismatches 142; Indels 7; Gaps 6;

QY 2 IQFSINRTLFIHAHTTKTAISTKNAIPILSSIKIEVTGTGVLTCNSGOISIENTIPVS 61
Db :|::||::||::||::||::||::||::||::||::||::||::||::||::||::||
1 MKFTIQKDLVESVQVDVLAVSSRTPILTGKIIVASDDGVSTGSDSDISIESFIPKE 60
QY 62 NENAGLL-TSPCAILELASFFNIITSSLPDISINVKIEHQHVLTSGSKSEITLKGDV 120
Db :|::||::||::||::||::||::||::||::||::||::||::||::||::||::
61 EGDKELVITEQPQSIVLOARFFSEIIVKKLFPMATVEIQVNQYLTIIRSGKAENFLNGDA 120
QY 121 DOYPRIQEYSTENPLTLTKLKSIIAETAFAASLOESRPIITGVHIVLSNHKDFKAVAT 180
Db :|::||::||::||::||::||::||::||::||::||::||::||::||::||::
121 DEYPHLPQIEEHAHQIPTDLLKNLRITQIVAVSTSETRPIITGVNWKV-EQSELLCTAT 179
QY 181 DSHRMSQRILTD-NTSADLMVVLPSPKSLREPSAFVETDDIETVEVPFPSPQLFRSEHIS 239
Db :|::||::||::||::||::||::||::||::||::||::||::||::||::||::
180 DSHRLRALRKALDIPEDRSYNNVVPKSLTELISKILDNDQELVDIVITETQVLFRAKNVL 239
QY 240 FYTRLEGNVPDTRLMTPEFTEVVFNTQSLRHAMERAFILSNATONGTVKLEI-TQNH 298
Db :|::||::||::||::||::||::||::||::||::||::||::||::||::||::
240 FFSRLLDGNYPDTSILPQDSKTETIIVTFKFLQIDRASLRAGEGRNVNWKSAKPES 299

```

QY	299	ISAHVNSPEVGVKVNEDLIVISQSGSGLTISFNPTYLISLKAIKSETVKIHLFSVPRPFT	350
Db	300	IEISSNSPEIGKVAVIADQTEGBELNISPSPKMLDALKVLEGAEIRVSTFGAMRPFL	359
QY	359	L-TPGDEEESFIQLITPVRT	377
Db	360	IRTFND--ETIVQLILPVRT	377
RESULT 33			
ID	Q81JD4	PRELIMINARY;	PRT; 381 AA.
AC	Q81JD4;		
DT	01-JUN-2003	(TrEMBLrel. 24, Created)	
DT	01-JUN-2003	(TrEMBLrel. 24, Last sequence update)	
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)	
DE	DNA polymerase III, beta chain (EC 2.7.7.7).		
GN	ORFNames=BC0002;		
OS	Bacillus cereus (strain ATCC 14579 / DSM 31).		
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.		
OX	NCBI_Taxid=226900;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
FX	MEDLINE=22608415; PubMed=12721630; DOI=10.1038/nature01582;		
RA	Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,		
RA	Kapatral V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,		
RA	Chu L., Mazur M., Goltseman E., Larsen N., D'Souza M., Walunas T.,		
RA	Grechkin Y., Fusch G., Haeelkorn R., Fonstein M., Ehrlich S.D.,		
RA	Overbeek R., Kyrpides N.C.;		
RT	"Genome sequence of Bacillus cereus and comparative analysis with		
RT	Bacillus anthracis.";		
RL	Nature 423:87-91(2003).		
DR	EMBL; AE016998; AAP07112.1; -		
DR	GO; GO:0008408; F:3'-5' exonuclease activity; IEA.		
DR	GO; GO:0003677; F:DNA binding; IEA.		
DR	GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.		
DR	GO; GO:0016740; F:transferase activity; IEA.		
DR	GO; GO:0006260; F:DNA replication; IEA.		
DR	InterPro; IPR001001; DNA_polIII_beta.		
DR	Pfam; PF00712; DNA_pol3_beta_1.		
DR	Pfam; PF02767; DNA_pol3_beta_2; 1.		
DR	Pfam; PF02768; DNA_pol3_beta_3; 1.		
DR	SMART; SMO0480; POL3bc; 1.		
DR	TIGRFAMs; TIGR00663; dnan; 1.		
KW	Nucleotidyltransferase; Transferrase.		
SEQ	SEQUENCE 381 AA; 42830 MW; 88F88D4710671163 CRC64;		
Query Match 36.58; Score 680; DB 2; Length 381;			
Best Local Similarity 38.44; Pred. No. 6.5e-37;			
Matches 147; Conservative 84; Mismatches 142; Indels 10; Gaps 8;			
QY	2	IQFSINRLPIHALNTTKRAISTKNAIPILSSIKIEVTSTGVLITGSGOISTENTIPVS	61
Db	1	MRFTIQDYLVRGVQDVKAVSSRTIPILTGKIVVATEEGVLITGSDADISTESFIPVE	60
QY	62	NEVAGLL-ITSPGAILLEASFNNISSLPDISINVKIEIQH-QVVLTSKSEITLKGKD	119
Db	61	ENGKEIVEITAQSGSIVLOAKYFSEIVKPLPKQVVEI-SVENHFWTKIKSGSEFNLGND	119
QY	120	VDOYPRLOEVSTENPLILTKLLKSIETAETAFASLQESRPIITGVHIVLSNHKDFKAVA	179
Db	120	SAEYPLLPQIEEHVHFVPIPTDLLKHMIRQTVFAVSTSETRPILITGVNWKYN-SELTGIA	178
QY	180	TDSHRMSQRLITLD-NTSAD---LMVVLPSKSLUREFSAVPTDDIETVEVFPSPQILFRS	235
Db	179	TDSHRLALRAKAKTEGNIHADEFOANVVIKPSKLNELSKILDESEEMVDIVITEQVQLFRT	238
QY	236	EHISFYTRLEGNYPTDRLILMTTEFEVVFNTQSLRHAMERAFILSNATQNGTVKLE-I	294
Db	239	KHLFFSRLEGNYPTDRLIPAESKTDIFVNTKEFLQADIRASLARDGRNNVVKLSTL	298
QY	295	TQNHISAHVNSPEVGVKVNEDLIVISQSGDLTISFNPTYLIESLKAIKSETVKIHLFSV	354

```

AC QRCAL;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE DNA polymerase III, beta chain (EC 2.7.7.7).
GN Name=dnan; OrderedLocuNames=BH0002;
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=99356711; PubMed=10427704;
RA Takami H., Masui N., Nakasone K., Horikoshi K.;
RT "Replication origin region of the chromosome of alkaliphilic Bacillus
RL halodurans C-125";
RL Biosci. Biotechnol. Biochem. 63:1134-1137(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RL halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- FUNCTION: DNA polymerase III is a complex, multichain enzyme
CC responsible for most of the replicative synthesis in bacteria.
CC This DNA polymerase also exhibits 3' to 5' exonuclease activity.
CC The beta chain is required for initiation of replication once it
CC is clamped onto DNA, it slides freely (bidirectional and ATP-
CC independent) along duplex DNA (By similarity).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA} (N).
CC -1- SUBUNIT: Contains a core (composed of alpha, epsilon, and theta
CC chains) that can repair short gaps created by nuclease in duplex
CC DNA. For efficient replication of the long, single-stranded
CC templates, pol III requires the auxiliary chains beta, gamma, and
CC delta (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on it
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. See http://www.isb-sib.ch/announcement
CC or send an email to license@sb-sib.ch.
CC
CC EMBL; AB013492; BAA82686.1; -
CC EMBL; AP001507; BAB03721.1; -
CC PIR; B83650; B83650.
CC InterPro; IPR001001; DNA_polIII_beta.
CC Pfam; PF00712; DNA_pol3_beta.1.
CC Pfam; PF02767; DNA_pol3_beta.2.1.
CC Pfam; PF02768; DNA_pol3_beta_3.1.
CC SMART; SM00480; POL3BC; 1.
CC TIGRFRAMS; TIGR00663; dnan; 1.
KW Complete proteome; DNA replication; DNA-directed DNA polymerase;
KW Transference.
SQ SEQUENCE 380 AA; 42382 MW; CA72B4944E115813 CRC64;
Query Match 33.5%; Score 624.5; DB 1; Length 380;
Best Local Similarity 35.2%; Pred. No. 3.1e-33;
Matches 134; Conservative 85; Mismatches 155; Indels 7; Gaps 5
QY 2 IQFSINRTLFTHALNTTKRAISTKNAIPILSSIKIEVSTGVTLTGSNGQISIENTIPV- 60
DB 1 MHFVIDRDFQNVNHSVKAVSRTTPILTGKIVADHEGVTLTGSQSDISIEFTIPLE 60
QY 61 SNENAGLLITSPGAILLEASPFNIIISLPLDISINVKIEHQVLTSGKSIITLKGKDV 120

```



```

Db 61 EGRQNVKQEGSVLQAKVFAEIVKQLPEQIEIHQVDSFVTTIRSGSVNENGLDP 120
QY 121 DQYPRQEVSTENPLILKTKLAKSIIAETAFASLOESRPLITGVHIVLNHDKDFKAVAT 180
Db 121 DEYPRPLVLEEDHVFRLPKILKIIIRQTVFAVSTQETRPVLTGVNFEED-GILTCTAT 179
QY 181 DSHRMSORLITLNDTSGADLM--VVLPSKSLRFSFAVFTDDIETVVFSPSPQILFRSH 237
Db 180 DSHRLAMRKVPVEKNDELQFSNVIPGKSLNELSKILDENBELLDIVVTENQTLFKLKN 239
QY 238 ISFTYTLLEGNYPDTDRLLMTETETEVNTQSLRHAMERAFILSNATQNTVKLEI-TQNH 299
Db 240 MLFFSRLLEGKYVPTKQMIPEKAKTSFAVHTAFLOTLERALLSREGKQNVINLKTLD 299
QY 297 NHTSAHVNSPEVGKVNEDDIVSQSGDLTIFSNPTLYLESKAIKSETVKIHFSPVRP 356
Db 300 GVVEVTAITPEIGKVTENATQGLEGEELRISFNGKNVIDALKVDSSEIHIAFTGAMSP 359
QY 357 FYLTPGDEESFQITLTPVRT 377
Db 360 FVLSPDTHDQS-LHLFSPVRT 379

RESULT 36
Q6HI40 PRELIMINARY; PRT; 376 AA.
ID Q6HI40
AC Q6HI40
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DE DNA polymerase III, beta subunit (EC 2.7.7.7).
GN Name=dnan; ORNNames=BT927.2460;
OS Bacillus thuringiensis serovar konkukian str. 97-27.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OX NCBI_TaxID=281309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=97-27;
RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017355; AAT61374.1; --
DR InterPro; IPR001001; DNA_polIII_beta.
DR Pfam; PF00712; DNA_pol3_beta; 1.
DR Pfam; PF02767; DNA_pol3_beta_2; 1.
DR Pfam; PF02768; DNA_pol3_beta_3; 1.
DR SMART; SM00480; POL3Bc; 1.
DR TIGRFAMs; TIGR00663; dnan; 1.
KW Nucleotidyltransferase; Transferase.
SQ SEQUENCE 376 AA; 41592 MW; E098AA29FA486A42 CRC64;

```

```

Query Match 30.6%; Score 569.5; DB 2; Length 376;
Best Local Similarity 34.4%; Pred. No. 1.3e-29;
Matches 130; Conservative 80; Mismatches 161; Indels 7; Gaps 7;

QY 2 IQFSINRFLFHALNTTKRAITSTNAIPILSSIKIEVTSTGVTLTGNGOISIENTIP-V 60
Db 1 MEFTVNHKHFTQALSEVSKAISTRAVILSGIKITADQSGITLIASNSNIFIEKFI 60
QY 61 SNEAGLLITSPGAILLEASFINTIISLPDISINVKIEHQHVLTSGKSEITLKKDV 120
Db 61 IDDSQITILQAGTIVVPAKFIIEIKKMP5-DIVIKSKNEQITIOSGITLNLGPPA 119
QY 121 DQYPRQEVSTENPLILKTKLAKSIIAETAFASLOESRPLITGVHIVLNHDKDFKAVAT 180
Db 120 NEFENVPLQDGHTEIQIETKQLIDAFKQTVFAVAKNESRPVLGTGVHIEL-DHKLICAA 178
QY 181 DSHRMSORLITLNDTSGADLMVVLPSKSLRFSFAVFTDDIETVVFSPSPQILFRSH 240
Db 179 DSHRLAIR-ETLITNNKANCIVSPSATTINELLKLMNSNLEFVSILYESHIIFTFGTTL 237

```

```

QY 241 YTRLEGNYPDTDRLLMTETETEVNTQSLRHAMERAFILSNATQNTVKLEI-TQNH 299
Db 238 YSRLEGYKPNISTILPNEFQTVINIDRQRMLOGVDRSSLLASEWANNVNLIVNESTI 297
QY 300 SAHVNSPEVGKVNEDDIVSQSG--SDLTIFSNPTLYLESKAIKSETVKIHFSPVRP 358
Db 298 QISSNASQIGIKSTQIQIDLGQEKQLNIFDGRFVMEVLKAIKEETITLTFGGSMRPL 357
QY 359 LTPGDEESFQITLTPVR 376
Db 358 IEAG-EQSAVVHLISPVR 374

RESULT 37
Q81PWO PRELIMINARY; PRT; 376 AA.
ID Q81PWO; Q6HY29; Q6KS44;
AC Q81PWO; Q6HY29; Q6KS44;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE DNA polymerase III, beta subunit.
GN Name=dnan-2; OrderedLocusNames=BA2684, BAS2499; ORNNames=GBAA2684;
OS Bacillus anthracis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames / isolate Porton;
RX MEDLINE=22608414; PubMed=12712629; DOI=10.1038/nature01586;
RA Read T.D., Peterson S.N., Tourasse N.J., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Ristone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.L.,
RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R.T., Redmond C., Thwaite J.E., White O.,
RA Salzberg S.L., Thomson B., Friedlander A.M., Kohler T.M.,
RA Hanna P.C., Kolstoe A.-B., Fraser C.M.;
RA "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria."
RL Nature 423:81-86(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ames / isolate 0581;
RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,
RA Fraser C.M.;
RT "Bacillus anthracis comparative genomics."
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Stearne;
RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Richardson P., Rubin E., Tice H.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017032; AAP26527.1; --
DR EMBL; AE017334; AAT31800.1; --
DR EMBL; AE017225; AAT54810.1; --
DR TIGR; BA2684; --
DR GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0004260; P:DNA replication; IEA.
DR InterPro; IPR001001; DNA_polIII_beta.
DR Pfam; PF00712; DNA_pol3_beta; 1.
DR Pfam; PF02767; DNA_pol3_beta_2; 1.
DR Pfam; PF02768; DNA_pol3_beta_3; 1.
DR SMART; SM00480; POL3Bc; 1.
DR TIGRFAMs; TIGR00663; dnan; 1.
SQ SEQUENCE 376 AA; 41705 MW; E57D2977F75E96F4E CRC64;

```



```
Query Match      29.9%; Score 557.5; DB 2; Length 376;
Best Local Similarity 33.9%; Pred. No. 8.4e-29;
Matches 128; Conservative 83; Mismatches 160; Indels 7; Gaps 7;

QY 2 IQFSINRTLFTHALNTTKRAISTKNAIPILSIKIEVTSTGVTLTGSGQISIENTIPVS 61
DB 1 MEFIVNHKHFQALSDVSKAISTKAIIPILSGIKITADQSGITLIASNSNPIEFKIPSA 60
QY 62 NENAGL-LITSPGAILLEASFFINIISLPDISINVKIEHQVVLTSKSEITLKGKDV 120
DB 61 IDEQITILQAGTIVVPKAFYFIEIIKKWPS-DIVIKSKNEQITIQSGEITLNLGPPA 119
QY 121 DOYPRLOEVSTENPLILTKLKSIIAETAFAASLOESRPILTGVHIVLSNHKOPKAVAT 180
DB 120 NEFPNVPQIDHTEIQIETKQIDAFKQTVFAVAKNESRHVLTGVHIEL-DHNKLI CAAT 178
QY 181 DSHRMSORLITLNTSADLMVLPSKSLRSESAVFTDDIETVEVFFSPQILFRSEHISF 240
DB 179 DSHRLAIR-ETLISTNMKANCIVPSATINELLKMNLSNLEFVSIYLSSESHIIFTGTTTL 237
QY 241 YTRLLEGNYPDTDRLLMTTEFEVFNFTQSLRHAMERAFILSNATQNGTVKLEI-TQNH 299
DB 238 YSRLIEGKYPNISTLIPINEFQTVINIDRQMLQGVDRSSLLASEWANNVNLVINESTI 297
QY 300 SAHVNSPEVGKVNEDLDI-VSQSGSDLTISFNPTVYLIESLKAISKSETVKIHLSPVRPFT 358
DB 298 QISSNASQIGKISEKQIDVIOGKQKLNISPDGRFMDALRAIKETVTLTSFGSGMRPIL 357
QY 359 LTPGDEEESFQILTPVR 376
DB 358 IEAGTQSAA-IHLISPV 374

RESULT 38
AAT31800
ID AAT31800 PRELIMINARY; PRT; 376 AA.
AC AAT31800;
DT 01-JUN-2004 (TREMBLrel. 27, Created)
DT 01-JUN-2004 (TREMBLrel. 27, Last sequence update)
DE DNA polymerase III, beta subunit.
GN DNAN-2 OR GBAA2684.
OS Bacillus anthracis str. Ames 0581.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus cereus group; Bacillus anthracis.
OX NCBI_TaxID=261594;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Ames 0581;
RA Ravel J., Raeko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S., Fraser C.M.;
RT "Bacillus anthracis comparative genomics.";
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017334; AAT31800.1; -.
DR EMBL; AE017334; AAT31800.1; -.
SQ SEQUENCE 376 AA; 41705 MW; E57D297F75E96F4E CRC64;

Query Match      29.9%; Score 557.5; DB 2; Length 376;
Best Local Similarity 33.9%; Pred. No. 8.4e-29;
Matches 128; Conservative 83; Mismatches 160; Indels 7; Gaps 7;

QY 2 IQFSINRTLFTHALNTTKRAISTKNAIPILSIKIEVTSTGVTLTGSGQISIENTIPVS 61
DB 1 MEFIVNHKHFQALSDVSKAISTKAIIPILSGIKITADQSGITLIASNSNPIEFKIPSA 60
QY 62 NENAGL-LITSPGAILLEASFFINIISLPDISINVKIEHQVVLTSKSEITLKGKDV 120
DB 61 IDEQITILQAGTIVVPKAFYFIEIIKKWPS-DIVIKSKNEQITIQSGEITLNLGPPA 119
QY 121 DOYPRLOEVSTENPLILTKLKSIIAETAFAASLOESRPILTGVHIVLSNHKOPKAVAT 180
DB 120 NEFPNVPQIDHTEIQIETKQIDAFKQTVFAVAKNESRHVLTGVHIEL-DHNKLI CAAT 178
QY 181 DSHRMSORLITLNTSADLMVLPSKSLRSESAVFTDDIETVEVFFSPQILFRSEHISF 240
DB 179 DSHRLAIR-ETLISTNMKANCIVPSATINELLKMNLSNLEFVSIYLSSESHIIFTGTTTL 237
QY 241 YTRLLEGNYPDTDRLLMTTEFEVFNFTQSLRHAMERAFILSNATQNGTVKLEI-TQNH 299
DB 238 YSRLIEGKYPNISTLIPINEFQTVINIDRQMLQGVDRSSLLASEWANNVNLVINESTI 297
QY 300 SAHVNSPEVGKVNEDLDI-VSQSGSDLTISFNPTVYLIESLKAISKSETVKIHLSPVRPFT 358
DB 298 QISSNASQIGKISEKQIDVIOGKQKLNISPDGRFMDALRAIKETVTLTSFGSGMRPIL 357
QY 359 LTPGDEEESFQILTPVR 376
DB 358 IEAGTQSAA-IHLISPV 374

RESULT 38
AAT31800
ID AAT31800 PRELIMINARY; PRT; 376 AA.
AC AAT31800;
DT 01-JUN-2004 (TREMBLrel. 27, Created)
DT 01-JUN-2004 (TREMBLrel. 27, Last sequence update)
DE DNA polymerase III, beta subunit.
GN DNAN-2 OR GBAA2684.
OS Bacillus anthracis str. Ames 0581.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;
OC Bacillus cereus group; Bacillus anthracis.
OX NCBI_TaxID=261594;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Ames 0581;
RA Ravel J., Raeko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,
RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S., Fraser C.M.;
RT "Bacillus anthracis comparative genomics.";
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017334; AAT31800.1; -.
DR EMBL; AE017334; AAT31800.1; -.
SQ SEQUENCE 376 AA; 41705 MW; E57D297F75E96F4E CRC64;

Query Match      29.7%; Score 553.5; DB 2; Length 376;
Best Local Similarity 33.9%; Pred. No. 1.5e-28;
Matches 128; Conservative 85; Mismatches 158; Indels 7; Gaps 7;

QY 2 IQFSINRTLFTHALNTTKRAISTKNAIPILSIKIEVTSTGVTLTGSGQISIENTIPVS 61
DB 1 MEFIVNHKHFQALSDVSKAISTKAIIPILSGIKITADQSGITLIASNSNPIEFKIPSA 60
QY 62 NENAGL-LITSPGAILLEASFFINIISLPDISINVKIEHQVVLTSKSEITLKGKDV 120
DB 61 IDEQITILQAGTIVVPKAFYFIEIIKKWPS-DIVIKSKNEQITIQSGEITLNLGPPA 119
QY 121 DOYPRLOEVSTENPLILTKLKSIIAETAFAASLOESRPILTGVHIVLSNHKOPKAVAT 180
DB 120 NEFPNVPQIDHTEIQIETKQIDAFKQTVFAVAKNESRHVLTGVHIEL-SNNK-LICAAT 178
QY 181 DSHRMSORLITLNTSADLMVLPSKSLRSESAVFTDDIETVEVFFSPQILFRSEHISF 240
DB 179 DSHRLAIR-ETLISTNMKANCIVPSATINELLKMNLSNLEFVSIYLSSESHIIFTGTTTL 237
QY 241 YTRLLEGNYPDTDRLLMTTEFEVFNFTQSLRHAMERAFILSNATQNGTVKLEI-TQNH 299
DB 238 YSRLIEGKYPNISTLIPINEFQTVINIDRQMLQGVDRSSLLASEWANNVNLVINESTI 297
QY 300 SAHVNSPEVGKVNEDLDI-VSQSGSDLTISFNPTVYLIESLKAISKSETVKIHLSPVRPFT 358
DB 298 QISSNASQIGKISEKQIDVIOGKQKLNISPDGRFMDALRAIKETVTLTSFGSGMRPIL 357
QY 359 LTPGDEEESFQILTPVR 376
DB 358 IEAGTQSAA-IHLISPV 374

RESULT 39
Q81C04
ID Q81C04 PRELIMINARY; PRT; 376 AA.
AC Q81C04;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE DNA polymerase III, beta chain (EC 2.7.7.7).
GN ORENAMES-BC2693;
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630; DOI=10.1038/nature01582;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapatal V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goltzman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Pusch G., Hagelkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyrpides N.C.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis.";
RL Nature 423:87-91(2003).
DR EMBL; AE017006; AAP09649.1; -.
DR GO; GO:0008408; F:3'-5' exonuclease activity; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006260; P:DNA replication; IEA.
DR InterPro; IPR001001; DNA_polIII_beta.
DR Pfam; PF00712; DNA_pol3_beta_1.
DR Pfam; PF02767; DNA_pol3_beta_2_1.
DR Pfam; PF02768; DNA_pol3_beta_3_1.
DR SMART; SM00480; POL3BC; 1.
DR TIGRFAMs; TIGR00663; dnan; 1.
DR Nucleotide; ttransferase; Transferase.
KW Nucleotide; ttransferase; Transferase.
SQ SEQUENCE 376 AA; 41736 MW; 9E9DF4C4AD18623B CRC64;

Query Match      29.7%; Score 553.5; DB 2; Length 376;
Best Local Similarity 33.9%; Pred. No. 1.5e-28;
Matches 128; Conservative 85; Mismatches 158; Indels 7; Gaps 7;

QY 2 IQFSINRTLFTHALNTTKRAISTKNAIPILSIKIEVTSTGVTLTGSGQISIENTIPVS 61
DB 1 MEFIVNHKHFQALSDVSKAISTKAIIPILSGIKITADQSGITLIASNSNPIEFKIPSA 60
QY 62 NENAGL-LITSPGAILLEASFFINIISLPDISINVKIEHQVVLTSKSEITLKGKDV 120
DB 61 IDEQITILQAGTIVVPKAFYFIEIIKKWPS-DIVIKSKNEQITIQSGEITLNLGPPA 119
QY 121 DOYPRLOEVSTENPLILTKLKSIIAETAFAASLOESRPILTGVHIVLSNHKOPKAVAT 180
DB 120 NEFPNVPQIDHTEIQIETKQIDAFKQTVFAVAKNESRHVLTGVHIEL-SNNK-LICAAT 178
QY 181 DSHRMSORLITLNTSADLMVLPSKSLRSESAVFTDDIETVEVFFSPQILFRSEHISF 240
DB 179 DSHRLAIR-ETLISTNMKANCIVPSATINELLKMNLSNLEFVSIYLSSESHIIFTGTTTL 237
QY 241 YTRLLEGNYPDTDRLLMTTEFEVFNFTQSLRHAMERAFILSNATQNGTVKLEI-TQNH 299
DB 238 YSRLIEGKYPNISTLIPINEFQTVINIDRQMLQGVDRSSLLASEWANNVNLVINESTI 297
QY 300 SAHVNSPEVGKVNEDLDI-VSQSGSDLTISFNPTVYLIESLKAISKSETVKIHLSPVRPFT 358
DB 298 QISSNASQIGKISEKQIDVIOGKQKLNISPDGRFMDALRAIKETVTLTSFGSGMRPIL 357
QY 359 LTPGDEEESFQILTPVR 376
DB 358 IEAGTQSAA-IHLISPV 374
```

Db 179 DSHRLAIR-ETLLSSDVKANCIVPSATINELLKMNNSBFFVYIYFSESHIIFTFTGTTTL 237  
Qy 241 YTRLLEGNYPDTORLLMTRETEVETVNTQSLRHAMERAFILSNATONGTVKLE-ITQNH 299  
Db 238 YSLIEGKYPNINLIPDNFKTIINVRKILQGVDRSSLASEWANNVNLINESTI 297  
Qy 300 SAHVNSPEVGVKNEDLDI-VSQSGDLTISFNPTYLIESLKAIKSETVKIHLSPVRPPT 358  
Db 298 KISSNASQIGKISSETQOIDAIOCEKQLNISFDGRFMDALRAIKETITLSFGGSNRPI 357  
Qy 359 LTPGDEESFIQITPVR 376  
Db 358 IEAG-EQSAVAHLISPV 374

## RESULT 40

Q737DO PRELIMINARY; PRT; 376 AA.  
AC Q737DO;  
DT 05-JUL-2004 (Tremblrel. 27, Created)  
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE DNA polymerase III, beta subunit (EC 2.7.7.7).  
GN Name-dnaN; OrderedlocusNames=BCE2720;  
OS Bacillus cereus (strain ATCC 10987).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=222523;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14960714;  
RA Rasko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,  
Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F.,  
Nelson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.,  
RT "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic  
adaptations and a large plasmid related to Bacillus anthracis pX01.",  
RL Nucleic Acids Res. 32:977-988(2004).  
DR EMBL; AE017272; RAS41632.1; -.  
DR TIGR; BCE2720; -.  
DR GO; GO:0003887; F:DNA-directed DNA polymerase activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR InterPro; IPR001001; DNA polIII beta.  
DR Pfam; PF00712; DNA\_pol3\_beta; 1.  
DR Pfam; PF02767; DNA\_pol3\_beta\_2; 1.  
DR Pfam; PF02768; DNA\_pol3\_beta\_3; 1.  
DR SMART; SM00480; POL3bc; 1.  
DR TIGRfams; TIGR00663; dnan; 1.  
SQ Complete proteome; Nucleotidyltransferase; Transferase.  
SQ SEQUENCE 376 AA; 41801 MW; ED9AFBCF7E5CBE31 CRC64;

Query Match 29.0%; Score 539.5; DB 2; Length 376;  
Best Local Similarity 33.3%; Pred. No. 1.3e-27;  
Matches 126; Conservative 82; Mismatches 163; Indels 7; Gaps 7;  
Qy 2 1QFSINRRLFIAHNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGNSQISIENTIPVS 61  
Db 1 MEFIWNKHGHTQALSEVSKALSTKAFIPILSGIKITANQSGITLIANSNIFIEKFPSS 60  
Qy 62 NENAGLLIT-SFGAILLEASPFINISSLDPDISINVKIEQHVVLTSKSEITLKGKV 120  
Db 61 MDDEQITTTLOAGTIVPAKYFIEIKKMP-S-DIVIKSKNEQITIQSEETLNLGPPA 119  
Qy 121 DOYPRLOEVSTENPLIKTLKLTIAETAASLQESRRPILTCGVHVLNHNKDFKAVAT 180  
Db 120 SEFNVPOIDHAEIQIETKQIDAFQTVFAVAKNESRPVLTCGVHIELDYNK-LICAA 178  
Qy 181 DSHRMSQRLITLDNTSADLMVLPKSLRPFSAVFTDDIETVEVFFSPQILFRSEHISF 240  
Db 179 DSHRLAIR-ETQISTNMKANCIVPSATINELLKMNNSLEFVSIVLSHSHIIFTFTGTT 237  
Qy 241 YTRLLEGNYPDTORLLMTRETEVETVNTQSLRHAMERAFILSNATONGTVKLEI-TQNH 299  
Db 238 YSLIEGKYPNINLIPDNFKTIINVRKILQGVDRSSLASEWANNVNLINESTI 297

Qy 300 SAHVNSPEVGVKNEDLDIVSQSG-DLTIISFNPTYLIESLKAIKSETVKIHLSPVRPPT 358  
Db 298 KISSNASQIGKISSETQOIDAIOCEKQLNISFDGRFMDALRAIKETITLSFGGSNRPI 357  
Qy 359 LTPGDEESFIQITPVR 376  
Db 358 IEAGTQSA-ITLISPV 374

Search completed: January 28, 2005, 16:59:36  
Job time : 76 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 16:50:32 ; Search time 41 Seconds

(without alignments)  
887.072 Million cell updates/sec

Title: US-10-048-071-28

Perfect score: 1863

Sequence: 1 MIQSIINRTLFHAIHTTKR.....LTPGDEBESFIQLTPVRTN 378

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1389	74.6	378	2 B95000	DNA polymerase III
2	1384	74.3	378	2 B97872	probable DNA-direc
3	1053	56.5	380	2 B86625	DNA-directed DNA p
4	1052	56.5	380	2 T30306	DNA-directed DNA p
5	730.5	39.2	377	1 S54708	DNA-directed DNA p
6	710	38.1	381	2 AC1433	DNA polymerase III
7	703	37.7	381	2 AC1432	DNA polymerase III
8	680.5	36.5	378	1 B22930	DNA-directed DNA p
9	624.5	33.5	380	2 B83650	DNA polymerase III
10	457.5	24.6	366	2 B96900	DNA polymerase III
11	423	22.7	363	1 S35733	DNA-directed DNA p
12	396.5	21.3	397	2 S70987	dnan protein - Myc
13	364	19.5	402	2 F70850	DNA-directed DNA p
14	362.5	19.5	367	1 DJPS3P	DNA-directed DNA p
15	354	19.0	366	1 A64107	DNA-directed DNA p
16	352	18.9	366	2 H82859	DNA polymerase III
17	350.5	18.8	367	2 F83644	DNA polymerase III
18	344	18.5	379	2 G97772	DNA-directed DNA p
19	339	18.2	366	2 F82376	DNA polymerase III
20	336.5	18.1	391	2 T47420	DNA-directed DNA p
21	335.5	18.0	399	2 T10002	DNA-directed DNA p
22	332	17.8	366	2 AB0958	DNA polymerase III
23	325	17.4	372	2 AE2613	DNA polymerase III
24	325	17.4	403	2 C97395	DNA polymerase III
25	323	17.3	381	2 B71700	DNA polymerase III
26	319.5	17.1	366	2 E72400	DNA polymerase III
27	318	17.1	366	1 DJBC3B	DNA-directed DNA p
28	318	17.1	366	2 D91208	DNA polymerase III
29	318	17.1	366	2 F86054	DNA polymerase III

30	317	17.0	366	2 AF0497	DNA-directed DNA p
31	315.5	16.9	456	2 AF2535	DNA polymerase III
32	314	16.9	387	2 AD2057	DNA polymerase III
33	314	16.9	397	2 AH3494	DNA-directed DNA p
34	307	16.5	367	1 JQ0734	DNA-directed DNA p
35	296.5	15.9	366	2 C81713	DNA polymerase III
36	296	15.9	376	2 B41870	DNA-directed DNA p
37	295.5	15.9	366	2 H72090	DNA polymerase III
38	295.5	15.9	366	2 B86533	DNA polymerase III
39	293.5	15.8	366	2 F81578	DNA polymerase III
40	292.5	15.7	416	2 E71559	probable DNA pol I
41	288	15.5	366	2 C84931	DNA-directed DNA p
42	287	15.4	372	2 C87268	DNA polymerase III
43	275	14.8	366	1 JC1159	DNA-directed DNA p
44	272.5	14.6	363	2 C70462	DNA polymerase III
45	270	14.5	367	2 C81030	DNA polymerase III

## ALIGNMENTS

### RESULT 1

B95000

DNA polymerase III, beta chain [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C>Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004

C:Accession: B95000

R:Tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei-

on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,

nson, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A>Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: B95000

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-378 <KUR>

A:Cross-references: UNIPROT:O06672; GB:AE005672; PIDN:AAK74195.1; PID:G14971466; GSPDB:GSPDB:G

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0002

C:Superfamily: DNA-directed DNA polymerase III beta chain

Query Match	74.6%	Score 1389;	DB 2;	Length 378;
Best Local Similarity	72.2%	Pred. NO. 1.6e-83;		
Matches 273;	Conservative 50;	Mismatches 55;	Indels 0;	Gaps 0;
Qy	1	MIQSIINRTLFHAIHTTKRAISTNAIPILSSIKIEVTSTCVTLTSGNGQISIENTIPV	60	
Db	1	MIHFSINKNLFQALNTTKRAISSKNAIPILSTVKIDVTNEGTTLIGNGQISIENTISQ	60	
Qy	61	SNENAGLLITSPGAILLEASFFINIISLPDISINVKIEHQVVLTSKSEITLKGKDV	120	
Db	61	KNEDAGLLITSGISLLEASFFINVVSSLPDVLDFKIEQNIQVLTSGKSEITLKGKDS	120	
Qy	121	DQYPLQVSTENPILTKLKSIIATFAASIQESRPILTGVIHVLNKHOFKAVAT	180	
Db	121	EQYPRIQEISASTPLILETKLKIINETAFASQESRPILTGVIHVLNKHOFKAVAT	180	
Qy	191	DSHRMSQRLITDNTSADLMVLPKSLRESASVTDDETVEVFPSQIILPRSEHISF	240	
Db	181	DSHRLSQKLTLEKNSDDFDVVPISRSREFSVFTDDIETVEIFANNQIILPRSENISF	240	
Qy	241	YTRLEGNYPDTDRLLMTTEFEVVFNTQSLRHAMERAFLLSNATQNGTVKLEITQNHS	300	
Db	241	YTRLEGNYPDTDRLLMTTEFEVVFNTQSLRHAMERAFLLSNATQNGTVKLEITQNHS	300	
Qy	301	AHVNSPEYGVKNEDLDIVSQSGSDLTISFNPTYLIESLKAISKETVKIHFSPVPPFTLT	360	
Db	301	AHVHSPGVKNEDLDIVSQSGSDLTISFNPTYLIESLKAISKETVKIHFSPVPPFTLT	360	
Qy	361	PGDEESFIQLTPVRTN	378	

C;Accession: B97872  
R;Kingsins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
Y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A;Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.  
A;Reference number: A97872; MUID:21429245; PMID:11544234  
A;Accession: B97872  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-378 <KUR>  
A;Cross-references: UNIPROT:O06672; GB:AE007317; PIDN:AAK98806.1; PID:gl5457530; GSPDB:G  
C;Genetics:  
A;Gene: dnan  
C;Superfamily: DNA-directed DNA polymerase III beta chain  
C;Keywords: nucleotidyltransferase

Query Match	74.3%;	Score 1384;	DB 2;	Length 378;
Best Local Similarity	72.2%;	Pred. No. 3.5e-83;		
Matches 273;	Conservative 49;	Mismatches 56;	Indels 0;	Gaps 0;
Qy	1	MIOFSNRTLFHIALNTTKRAISTKKAIPILSSIKTIEVSTGTVLTGNSQGISIENTIPV	60	
Db	1	MIHFSINKNLFQALNITTKRAISKKAIPILSTVKIDVTNEGVTILGNSQGISIENTISQ	60	
Qy	61	SNENAGLLITSPGAILLEASFFINIISSLPDISINVKIEIQHQQVLTSGKSEITLKGKV	120	
Db	61	KNEDAGLLITSLGSIILEASFFINVVSSLPDVLTDPKEIQNQIVLTSGKSEITLKGKDS	120	
Qy	121	DOYPRQEVSTENPLILKTKLLSIITAETAFAASLOESRPILTGVIHLVLSNHKDFKAVAT	180	
Db	121	EQYPRIQEISASPTLILKTKLLKKIINETAFAASTQESRPILTGVIHVPVLSQHKELTKTAV	180	
Qy	181	DSHRMSORLITLONTGADLMVPLPSKSLRFSFAVFTDDIETVEVFFSPSOILFRSEHISF	240	
Db	181	DSHRLSQKLTLEKNSDDFVPIPSRSLRFSFAVFTDDIETVEIFPANNQILFRSENISF	240	
Qy	241	YTRLLGEGYPTDRLMLMTEFETEVFNWNTQSLRHAMERAFILSNATONGTVKLEBITQNHIS	300	
Db	241	YTRLLGEGYPTDRLLIPTDPTNTITFNVNVNLRQSMERARLLSSATONGTVKLEIKDGWVS	300	
Qy	301	AHVNSPEVGKVNEDDLTVSOGSDLTISFNPTYLIESLKAISKETVKIHFSLSPVRPTLT	360	
Db	301	AHVHSPVGKVNIEDTQVGTGDLTISFNPTYLIDSLKALNSEKVTISFISAVRPTTLV	360	
Qy	361	PGDEESFIQLITPVRTN	378	
Db	361	PADTDQFMOLITPVRTN	378	

RESULT 3  
B86625  
DNA-directed DNA polymerase (EC 2.7.7.7) beta chain [imported] - Lactococcus lactis subsp. N; Alternate names: DNA polymerase III, beta chain  
C/Species: Lactococcus lactis subsp. lactis  
C/Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 09-Jul-2004  
C/Accession: B86625  
R/Bolotin, A.; Winkler, P.; Mauger, S.; Jaillon, O.; Mialme, K.; Weissenbach, J.; Ehrlich Genome Res. 11, 731-753, 2001  
A/Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. A; Reference number: A86625; MUID:21235186; PMID:11337471  
A/Accession: B86625  
A/Status: preliminary

RESULT 4  
T30306  
DNA-directed DNA polymerase (EC 2.7.7.7) III beta chain - Lactococcus lactis  
C/Species: Lactococcus lactis  
C/Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T30306  
R;El-Karoui, M.; Ehrlich, D.; Gruss, A.  
Proc. Natl. Acad. Sci. U.S.A. 95, 626-631, 1998  
A/Title: Identification of the lactococcal exonuclease/recombinase and its modulation by  
A/Reference number: Z20817; MUID:98118563; PMID:9435243  
A/Accession: T30306  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-380 <ELK>  
A/Cross-references: UNIPROT:O54376; EMBL:U76424; NID:G2909712; PID:G2909714; PIDN:AAC1291  
C/Genetics:  
A/Gene: dnan  
C/Superfamily: DNA-directed DNA polymerase III beta chain  
C/Keywords: nucleotidyltransferase

```
Query Match      56.5%; Score 1052; DB 2; Length 380;  
Best Local Similarity 54.7%; Pred. No. 1.8e-61;  
Matches 208; Conservative 74; Mismatches 96; Indels 2; Gaps 1;
```

---

```
Qy    1  MIQPSINRTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVLTGSGNQGISIENTIPV 60  
       |.:|.|.|.:|.:.|.|.:|.:.|.|.:|.:.|.|.:|.:.|.|.:|.:.|.|.:|.:.|.  
Db    1  MIRFSINKTAFQNALKITTKQAIGSKVTIPTALPKLKEVEEKGITLLGSNGQISIKNLP 60  
       |.:|.|.|.:|.:.|.|.:|.:.|.|.:|.:.|.|.:|.:.|.|.:|.:.|.|.:|.:  
Qy   61  SNEVAGILLTSPGAIIILEASPFNIISLDPIISNVKETEIQHOVVLTSGKSEITLKGDV 120  
       |.:|.|.|.:|.:.|.|.:|.:.|.|.:|.:.|.|.:|.:.|.|.:|.:.|.|.:|.:  
Db   61  DNKDASMLISGTGVLEAAFPFENVVSOLPEVTLFIEKEQKVOLLTSGKSEITLKGDLS 120
```

Query Match	39.2%;	Score	730.5;	DB	1;	Length	377;
Best Local Similarity	39.3%;	Pred. NO.	1.8e-40;				
Matches	149;	Conservative	90;	Mismatches	135;	Indels	5;
Gaps	4						
Qy	1	MIQFSINRTLPFIHANTTKRAISTKNALPIILSSIKIEVTSGTVTITGSGNQIISIENTIPV	60				
Db	1	MMEFTIKRDYFYTQNDLTALKALSFRPTTPIUTIGIKIDAKEHEVILUTGSDSEISIIITPK	60				
Qy	61	SNENAGLL-ITSPGAILLEASFFNNIIISLPDISINVKIEIQHQVVLVTSKGSEITLKGKD	119				
Db	61	TVDGEDIVNISETGSGVLPGRFFVDI IKKLPGKDVKLSTNEQFQTLITSGHSEFNLSGLD	120				
Qy	120	VDQYPRLOEVSTENPLIIKTKLLKSIIIAETAFASLOESRPLITGVHIVLSNHHKDFKAVA	179				
Db	121	PDQYPLLPQVRDDAIQUSVLKRVNIAQTFNAVSTSETRPLVTLGVNMLIOEN-ELICTA	179				
Qy	180	TDSHRMSORLITLTONTSADLMVLPKSLRBFSAVFTDDIETVEVFFSPSQILPSESHIS	239				
Db	180	TDSHRLAVRKLQLEDVSENKNVIPGKALAEINKMSDNEEDIDIFFASNOVLPKVGNNV	239				
Qy	240	PYTLLEGNYDDTRLLMTBETEVVFNQTSLRHAMERAF LISNATQNGTVKLEITONHI	299				
Db	240	FISRLLEGHYDDTRLPFPENYEIKLSDINGEFYHAIDRASLLAREGGNNVILKSTGDDVV	299				
Qy	300	SAHVNSPEVGKVNEDLDIVSOGSDLTISFNPTYLIESLKATKSTVKTHFLSPVRPFTL	359				
Db	300	ELSTSPETIGTVKEVDANDVEGSLKISFNKSNKMYMDALKALDNDEVEVEFFGTMKPFIL	359				
Qy	360	TP-GDEESFIQLITTPVRT	377				
Db	360	KPKGD--DSVTQLILPIRT	376				

RESULT 7  
AC1432  
DNA polymerase III, beta chain [imported] - *Listeria monocytogenes* (strain EGD-  
C) Species: *Listeria monocytogenes*  
C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C/Accession: AC1432  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.;  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.;  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournon,  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.;  
A>Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; PMID:21537279; PMID:11679669

A:Accession: AC1432  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-381 <GLA>  
A:Cross-references: UNIPROT:ORVAM1; GB:NC\_003210; PIDN:CAC98217.1; PID:gl6409361; GSPDB:  
A:Experimental source: strain EGB-e  
C:Genetics:  
A:Gene: dnaN  
C:Superfamily: DNA-directed DNA polymerase III beta chain

Query Match 37.7%; Score 703; DB 2; Length 381;  
Best Local Similarity 38.8%; Pred. No. 1.2e-38;  
Matches 148; Conservative 93; Mismatches 134; Indels 6; Gaps 5;

QY 2 IOFSINRTLFHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNGQISIENTIP-V 60  
Db 1 MKFVIERDLVQAVNEVTRAIASRTIPIITGKIIVNDEGVTLTGSDSDISIEAFIPI 60

QY 61 SHENAGLITSGAILLEASFINIISLPDISINNVKIEHQVLTGSKSEITLKGKDV 120  
Db 61 ENDEVIVEVESGGIVLOSKEYFGDIVRRLPENVEIETVTSNYQTNISSQASFTLGLDP 120

QY 121 DOYPRLOEVSTENPILKTLKLSIAETAPAAQLQESRPILTGVHIVLSNHKDFKAVAT 180  
Db 121 MEYPKLPEVTDGKTIPINVLKNIIVRQTVFAVSAIEVRPVLTVGNVLIKENK-LSAVAT 179

QY 181 DSHRMSORLITLD-NTSADLMVLPFSKSLREFSAVFTDDIETVEVFPSQILFRSEHIS 239  
Db 180 DSHRLALREIPIETDIDEBYNIVIPGKSLSELNKLDDASESIEMTLANNQILFKLKDLL 239

QY 240 FYTRLLEGNYPTDRLHMTTEFTEVFNQSLRHAMERAFILSNATONGTVKLEITQN-H 298  
Db 240 FYSLRLEGSYPDTSRLIETDTKSELVINSKAPLQADIRASLLARENNVNIKMLTLENGQ 299

QY 299 ISAHVNSPEVGVKNEDLDIVSOGSDLTISFNPTVLIRESLKAISKETVKIHFSPVRPT 358  
Db 300 VEVNSNSPEVGVNSVNSQSTGGEIKISFNKMGKTMADALRAFEQDDIQISFGTMRPFV 359

QY 359 LTPGD-BEESFIQITPVRT 377  
Db 360 LRPKDAANPNEILQITPVRT 380

RESULT 8  
B22930  
DNA-directed DNA polymerase (SC 2.7.7.7) III beta chain dnaN - Bacillus subtilis  
N:Alternate names: 42K protein (oriC region)  
C:Species: Bacillus subtilis  
C:Date: 21-May-1988 #sequence revision 16-Aug-1996 #text\_change 09-Jul-2004  
C:Accession: B22930; S66032; I40393; A69618  
R:Moriya, S.; Ogasawara, N.; Yoshikawa, H.  
Nucleic Acids Res. 13, 2251-2265, 1985  
A:Title: Structure and function of the region of the replication origin of the Bacillus  
A:Reference number: A94702; MUID:85215612; PMID:2987847  
A:Accession: B22930  
A:Molecule type: DNA  
A:Residues: 1-378 <OR>  
A:Cross-references: UNIPROT:P05649; GB:D26185; NID:G467326; PIDN:BAA05238.1; PID:G467392  
R:Ogasawara, N.; Moriya, S.; von Meyenburg, K.; Hansen, F.G.; Yoshikawa, H.  
EMBO J. 4, 3345-3350, 1985  
A:Title: Conservation of genes and their organization in the chromosomal replication ori  
A:Reference number: A26057; MUID:86135970; PMID:3004954  
A:Contents: annotation  
R:Ogasawara, N.; Nakai, S.; Yoshikawa, H.  
DNA Res. 1, 1-14, 1994  
A:Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chrom  
A:Reference number: S65967; MUID:96051385; PMID:7584024  
A:Accession: S66032  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-378 <OGA>  
A:Cross-references: EMBL:D26185; NID:G467326; PIDN:BAA05238.1; PID:G467392  
R:Moriya, S.; Fukuoka, T.; Ogasawara, N.; Yoshikawa, H.

EMBO J. 7, 2911-2917, 1988  
A:Title: Regulation of initiation of the chromosomal replication by DnaA-boxes in the ori  
A:Reference number: I40391; MUID:89030659; PMID:2846289  
A:Accession: I40393  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-30 <RES>  
A:Cross-references: EMBL:X12779; NID:G39878; PIDN:CAA31271.1; PID:G39879  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
C.; Bron, S.; Brouillett, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
A.; Ehrlich, S.D.; Emmerison, P.F.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; E  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koecker, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terptrala, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: A69618  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-378 <KUN>  
A:Cross-references: GB:Z99104; GB:AL009126; NID:G2632267; PIDN:CAB11778.1; PID:G2632269  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: dnaN  
A:Map position: 0  
C:Superfamily: DNA-directed DNA polymerase III beta chain  
C:Keywords: DNA replication initiation; nucleotidyltransferase

Query Match 36.5%; Score 680.5; DB 1; Length 378;  
Best Local Similarity 37.1%; Pred. No. 3.4e-37;  
Matches 141; Conservative 90; Mismatches 142; Indels 7; Gaps 6;

QY 2 IOFSINRTLFHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNGQISIENTIPVS 61  
Db 1 MKFTIQKDLVESQDVLVKAVSSRTTIPITGKIIVASDDGVSTGSDSDISIESFIPKE 60

QY 62 NENAGL-ITSPGAILLEASFINIISLPDISINNVKIEHQVLTGSKSEITLKGKDV 120  
Db 61 EGDKEIVTIEQPGSVIQLQARFSEIVKLPMTATVEIVQNYLTIIIRSGKAFFNLGLDA 120

QY 121 DOYPRLOEVSTENPILKTLKLSIAETAPAAQLQESRPILTGVHIVLSNHKDFKAVAT 180  
Db 121 DEYPHLPQIEBHAIQIPTDILLKNIIRQTVFAVSTSETRPILTGVNWKV-BQSELLCTAT 179

QY 181 DSHRMSORLITLD-NTSADLMVLPFSKSLREFSAVFTDDIETVEVFPSQILFRSEHIS 239  
Db 180 DSHRLALRKAKLDIPEDRSVNVIPGKSLTSLKILDDNDELVDIVITETQVLFKANKVL 239

QY 240 FYTRLLEGNYPTDRLHMTTEFTEVFNQSLRHAMERAFILSNATONGTVKLEI-TQNH 298  
Db 240 FFSRLLDGNYPTDTSLLIPQDSKTEIIVNTRKEFLQADIRASLLAREGNVNVKLSAKPAES 299

QY 299 ISAHVNSPEVGVKNEDLDIVSOGSDLTISFNPTVLIRESLKAISKETVKIHFSPVRPT 358  
Db 300 IEISSNSPEIGKVVEAIVADQIEGELNISFSPKPKYDALKVLEGAIRVFTGAMRPF 359

QY 359 LTPGD-BEESFIQITPVRT 377  
Db 360 IRTEND--ETIVQILLPVRT 377

RESULT 9  
B83650  
DNA polymerase III (beta subunit) dnaN [imported] - Bacillus halodurans (strain C-125)  
C:Species: Bacillus halodurans





R;Fujita, M.Q.; Yoshikawa, H.; Ogasawara, N.  
Mol. Gen. Genet. 215: 381-387; 1989.  
A;Title: Structure of the dnaA region of *Pseudomonas putida*: conservation among three  
A;Reference number: Jv0002; MUID:89218947; PMID:2540413  
A;Accession: Jv0002  
A;Molecule type: DNA  
A;Residues: 1-367 <FUJ>

RESULT 14  
DJPS3P  
DNA-directed DNA polymerase (EC 2.7.7.7) III beta chain - *Pseudomonas putida*  
C/Species: *Pseudomonas putida*  
C/Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004  
C/Accession: JTV0002

C; Date: 30-Jun-1992 #sequence=revision 30 Jun 1992 #name=

A; Accession: JV0002  
R; Fujita, M.Q.; Yoshikawa, H.; Ogasawara, N.  
Mol. Gen. Genet. 215, 381-387, 1989  
A; Title: Structure of the dnaA region of Pseudomonas putida: conservation among three b  
A; Reference number: JV0002; MUID:89218947; PMID:2540413  
A; Accession: JV0002  
A; Molecule type: DNA  
A; Residues: 1-367 <FUJ>





C;Superfamily: DNA-directed DNA polymerase III beta chain

Query Match 18.2%; Score 339; DB 2; Length 366;  
Best Local Similarity 23.2%; Pred. No. 7e-15;  
Matches 87; Conservative 94; Mismatches 184; Indels 10; Gaps 5;

QY 2 IQPSINRFLPHALNTTKRAISTKVAIPILSIKIEVTSTGVTLTGSGNGQISIENTIPVS 61  
DB 1 MKPTIERSHLIKPQQVSGTGLGGRASLPILGNLLKVEENQLSMTATDELISRVTL 60  
QY 62 NENAGLLITSPGAILLEASFFINIISSLPDTSINVKIEHQVLTSGKSEITLKGKVD 121  
DB 61 GE-----FEAGSIIVPARKELDICRGLPDSAVITVLEGBRIQVRSRFSLATLPAS 114  
QY 122 QYPRLOEVSTENPLILKLLKSIITAETAFAASLQESRPILTGVHIVLSNHKDFKAVATD 181  
DB 115 DFPNIEDQSEVQVSLTQAEIRGLIEKTFQSMANQDVRYLNGMLFEIDG-TTLRSVATD 173  
QY 182 SHRSQRLITLTDNTSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHISFY 241  
DB 174 GHRMAVAQAQGLADFAQKQIIVPRKGVLELVKLLDAPQPVVLIQIGHNLRAEVNHVFT 233  
QY 242 TRLEGNYPTDRLMTFETFEVNTQSLRHAMERAFILSNATQNGTVKLEITONHISA 301  
DB 234 SKLVGDRFPDYRVLPQHTSKTLQGCCELRAQFSAAILSNEXFRG-VRVNLADNGMRI 292  
QY 302 HVNSEVQKVNEDLDIVSQSGDLTISFNPTYLIESLKAISKSETVKIHLSPVRFTLTP 361  
DB 293 TANNPEQAEAEELLD-VSFEQEPFEGFVNSVILDVNLTRCDNVRVS-MSDANASALVE 350  
QY 362 GDEESFQLITPVR 376  
DB 351 NVDDSDAMVVMPIR 365

# RESULT 20

S74720  
DNA-directed DNA polymerase (EC 2.7.7.7) III beta chain - *Synechocystis* sp. (strain PCC N;Alternate names: protein slr0965

C;Species: *Synechocystis* sp.  
A;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C;Accession: S74720  
F;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* s.

A;Reference number: S74322; MUID:97061201; PMID:8905231

A;Accession: S74720

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-391 <KAN>

A;Cross-references: UNIPROT:P72856; EMBL:D90901; GB:AB001339; NID:G1651897; PIDN:BAAL687  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C;Genetics:

A;Gene: dnaN

C;Superfamily: DNA-directed DNA polymerase III beta chain

C;Keywords: DNA replication initiation; nucleotidyltransferase

Query Match 18.1%; Score 336.5; DB 2; Length 391;  
Best Local Similarity 26.0%; Pred. No. 1.1e-14;  
Matches 101; Conservative 79; Mismatches 168; Indels 41; Gaps 12;

QY 15 LNTTKRAISTKVAIPILSIKIEVTS--TGVTLTGSGNGQISIENTIPVSNENAGLLITSP 72  
DB 14 LSLVSRVSSRPTHPLVGNLLEADADKNYUKLTAFDLSLQSSFTAD-----VQGS 66  
QY 73 GAILLEASFFINIISSLPDISIN-----KEIEHQVLTSGKSEITLKGKVDQYPR 126  
DB 67 GRITLPKALNDIVSRDPDITLADPDGAGDGHLLTITSESGRFOIRGLDADDFAL 126  
QY 127 QEVSTENPLILKLLKSIITAETAFAASLQESRPILTGVHIVLSNHKDFKAVATDSHRMS 186

DB 127 PTVEGVKPLLPVATLNEGLRGAALFAASTDTKQVLTGVHIKSGG-DSLEFAATDGHRLA 185  
QY 187 -----ORLITLTDNTSADLMVLPKSLRSEPSAVFTD--DIETVEVFPSPQI 231  
DB 186 VVEAPTQIENDEGEAVIT-GSDLADPAVTIIPARALRELERMVASQSGNDLSLVNDVTQV 244  
QY 232 LFRSEHISFYTRLLEGNYPTDRLMTFETFEVNTQSLRHAMERAFILSNATQNGTV 290  
DB 245 IFELGDORLTSRKLEGAPAYDQLIPRQFRTVTMERKRLITSLERSVLADQKNLVTF 304  
QY 291 KLEITQNHISAHVNSPEVKVNEEL--DIVSQSGDLTISFNPTYLIESLKAISKSETVKI 348  
DB 305 TLQSPGNQLQAVSAQDLHGHEESMGAETIIEGGQ---IAFNIRKYLMDGLKALPNDIQM 361  
QY 349 HFLSPVRFTLTP-GDEESFQLITPVR 376  
DB 362 QLNENQPVITPTPLGGLKMTY--LVMPVR 388

# RESULT 21

TI10002

DNA-directed DNA polymerase (EC 2.7.7.7) III beta chain dnaN - *Mycobacterium leprae*  
C;Species: *Mycobacterium leprae*  
C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004  
C;Accession: TI10002

R;Cole, S.T.

submitted to the EMBL Data Library, August 1997

A;Reference number: Z16916

A;Accession: TI10002

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-399 <COL>

A;Cross-references: UNIPROT:P46387; EMBL:Z70722

C;Genetics:

A;Note: dnaN

C;Superfamily: DNA-directed DNA polymerase III beta chain

C;Keywords: nucleotidyltransferase

Query Match 18.0%; Score 335.5; DB 2; Length 399;  
Best Local Similarity 25.9%; Pred. No. 1.3e-14;  
Matches 104; Conservative 82; Mismatches 171; Indels 45; Gaps 11;

QY 2 IQPSINRFLPHALNTTKRAISTKVAIPILSIKIEVTSTGVTLTGSGNGQISIENTIPVS 61  
DB 13 LKPCLARESFASAVSWAKYLPTRPTVPVLSGVLITGSDSLTISGDFVEVSAE--VQVA 70  
QY 62 NENAGLLITSPGAILLEASFFINIISSLPDTSINVKIEHQVLTSGKSEITLKGKVD 121  
DB 71 AE-----IASSGSVLVSGRLSDITRALPNKPVHF-YVDGNRVALTGSGARFSLPTMAVE 124  
QY 122 QYPRLOEVSTENPLILKLLKSIITAETAFAASLQESRPILTGVHIVLSNHKDFKAVATD 181  
DB 125 DYPTLPTLPTDTC-TLSDVFAEAIQVAIAAGRDYTLPLMTGIRIEISGDTVVLA-ATD 182  
QY 182 SHRSQRLITLTDNTSADL--MVLPKSLRSEPSAVFTDDIETV-----EVFSP 228  
DB 183 RFLAVRELKWSVLSSDFEASVLPVPAKTLVEAKAGTDSGVCLSLGAGVGVGKDLFGI 242  
QY 229 SQILFRSEHISFYTRLLEGNYPTDRLMTFETFEVNTQSLRHAMERAFILSNATQNG 288  
DB 243 SGGGKRST-----TRLDAEFPKFRQLLPASHATAVATIDVAELTEAIKLVAV--ADRG 295  
QY 289 TVKLEITQNHISAHVNSPEVKVNEELDIVSQSGDLTISFNPTYLIESLKAISKSETVKI 348  
DB 296 QVRNEFGDGIURLSAGADDVGRAEEL--AVAFTESEPLTIANPNVLTGLASVHSRVSF 354  
QY 349 HFLSPVRFTLTPGDDEE-----SFQLITPVR 376  
DB 355 GFTTSPKALLRPTNSDDVHPTDHPGPPALPTDYVYLLMPVR 396

# RESULT 22

AB0958  
DNA polymerase III beta-chain [imported] - Salmonella enterica subsp. enterica serovar Typhimurium  
C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C:Accession: AB0958  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, P.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium  
A:Reference number: AB0502; MUID:21534947; PMID:11677608  
A:Accession: AB0958  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-366 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD03158.1; PID:g16504793; GSPDB:GN00176  
C:Genetics:  
C:Superfamily: DNA-directed DNA polymerase III beta chain

Query Match 17.8%; Score 332; DB 2; Length 366;  
Best Local Similarity 24.0%; Pred. No. 2e-14;  
Matches 90; Conservative 90; Mismatches 185; Indels 10; Gaps 5;

Qy 2 IQFSINRTLFTHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGNGOISIENTIPVS 61  
Db 1 MKFTVERHLKPKQVSGPLGGRTPLFGLNLLQVADGTLSTGTDLEMEMVARVTL 60

Qy 62 NENAGLLITSPGAILLEASFFINISSLIPDISINVKEIQHOVLTSGSEITLKGD 121  
Db 61 QPH-----EPGATVPARKFDCRGLPEGAETAIVQLGDMVRSGRSFSLTLPAA 114

Qy 122 QYPRLOEVSTENPLIKTLKLSIIAETAFAASQESRPILTGV--HIVLSNHK-DPK 181  
Db 115 DFPNLDWQSEVFTLPQATKRLIEATQFSMAHQDVRYILNGM-LFETEGSELRTVATD 173

Qy 182 SHRMSQRLITLNTSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHSFY 241  
Db 174 GRLIACVSMPLASLPSSHVSIVPRKGVIELMRMLDGGENPLRVQIGSNIRAHVGDFT 233

Qy 242 TRLEGNYPTDRLMTTEFEVFNTPQSLRHAMERAFILSNATQNGTVKLEITQNHLISA 301  
Db 234 SKLVDRFPDYRRVLPKPNPKHLEAGCDILQAAPARAAILSNKFRG-VRLVSENQLKI 292

Qy 302 HVNPEVKVNEEDDIVSQSGDLTISNPNYILSLKAIKSETVKIHFSPVFPFLTP 361  
Db 293 TANNPEQEAEEILD-VSYGTEMEIGFNVSIVLDVLNALKCTVRIMLTDSVSVQIED 351

Qy 362 GDEESFQLITPVR 376  
Db 352 A-ASQSAAYVMPMR 365

RESULT 23  
AE2613  
DNA polymerase III, beta chain [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 09-Jul-2004  
C:Accession: AE2613  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, K.; P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; MUID:21608550; PMID:11743193  
A:Accession: AE2613  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-372 <KUR>

A:Cross-references: UNIPROT:Q8UIJ4; GB:AB008688; PIDN:AAL41323.1; PID:g17738634; GSPDB:G  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: dnaN  
A:Map position: circular chromosome  
C:Superfamily: DNA-directed DNA polymerase III beta chain

Query Match 17.4%; Score 325; DB 2; Length 372;  
Best Local Similarity 25.5%; Pred. No. 5.9e-14;  
Matches 98; Conservative 86; Mismatches 178; Indels 22; Gaps 11;

Qy 2 IQFSINRTLFTHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGNGOISIENTIPVS 61  
Db 1 MRITLERSNLLKSLNHVVRVRRNTPIILSNVLLRSGANLDMKATDLDEITEATPAM 60

Qy 62 NENAGLLITSPGAILLEASFFINISSLIPDISINVKEIQHOVLTSGSEITLKGD 119  
Db 61 VEQAG-ATTVPALHLYE-----IVRKLDPGSEVILLATNPDGSSMTVASGRKFSLOCLP 113

Qy 120 VDQYPRLOEVSTENPLIKTLKLSIIAETAFAASQESRPILTGV--HIVLSNHK-DPK 176  
Db 114 EADFPDLITAGTFSHTFKLKAADLKWLDRTQFAISTETRYVYLANGIFPHITIESNGELKLR 173

Qy 177 AVATDSHRMQRLLTDLNTSADLM-VVLPKSLREFSAVFTDDIETVEVFPSPQILPVS 235  
Db 174 AVATDGHRLARADVADPSGSEGMPIIPRKTGVLQKMDNPELVTVEVSDAKIRLAI 233

Qy 236 EHISFYTRLEGNYPDTDRLLMTTEFEVFNTPQSLRHAMERAFILSNATQNGTVKLEIT 295  
Db 234 GSVVLTSLKLDGTQYQYRVIPTGNDKEMRVDCQTFARADVRSVTIS-SERGRVVKALT 292

Qy 296 QNHISAVNSPEVKVNEEDDIVSQSGDLTISNPNYILSLKAIKSETVKIHFSPVR 355  
Db 293 DGQTLTVNPPDSSGATEEV-AGVDNDSMEIGNAKYLLDITSLQSGEDA-IFLLADAG 350

Qy 356 PFTL---TPGDEESFQLITPVR 376  
Db 351 SPTLVDRDTAGDA---LYVLMPMR 371

RESULT 24  
C97395  
DNA polymerase III, beta chain [imported] - Agrobacterium tumefaciens (strain C58, Cerezo)  
C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 09-Jul-2004  
C:Accession: C97395  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens C58  
A:Reference number: A97359; MUID:21608551; PMID:11743194  
A:Accession: C97395  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-403 <KUR>  
A:Cross-references: UNIPROT:Q8UIJ4; GB:AE007869; PIDN:AAK86116.1; PID:g15155199; GSPDB:G  
C:Genetics:  
A:Gene: AGR C 520  
A:Map position: circular chromosome  
C:Superfamily: DNA-directed DNA polymerase III beta chain

Query Match 17.4%; Score 325; DB 2; Length 403;  
Best Local Similarity 25.5%; Pred. No. 6.6e-14;  
Matches 98; Conservative 86; Mismatches 178; Indels 22; Gaps 11;

Qy 2 IQFSINRTLFTHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGNGOISIENTIPVS 61  
Db 32 MRITLERSNLLKSLNHVVRVRRNTPIILSNVLLRSGANLDMKATDLDEITEATPAM 91

Qy 62 NENAGLLITSPGAILLEASFFINISSLIPDISINVKEIQHOVLTSGSEITLKGD 119  
Db 92 VEQAG-ATTVPALHLYE-----IVRKLDPGSEVILLATNPDGSSMTVASGRKFSLOCLP 144

```
QY 120 VDOYPRLOEVSTENPLILKTKLLKSLIAETAFASLOESRPILTGV--HIVLSNHK-DFK 176
Db 145 EADFDELDTAGTSHFTKKAADLKMILDRTOFAITESTETRYYLNGIFFHTIESNGELKRL 204
QY 177 AVATDSHRMSQRLITLIDNTSADLM-VVLPSPKSLREFSAVFTDDIETVEFFSPSOILFRS 235
Db 205 AVATDGHRLARADVDPAPSGSEGMPIIIPRKTVGELQKMDNPELEVTVESDAKIRLAI 264
QY 236 EHISFYTRLEGNYPDRLMTETETEVVFNQSLRHAMERAFILSNATQNGTVKLEIT 295
Db 265 GSVVLTSKLIDGTFDQYRVPTGNDKEMRVDCQTFARAVDRVSTIS-SERGRAVKLAIT 323
QY 296 QNHISAHVNSPVGKVNEDLDIVSQSGDLTISFNPTVILIESLKAIKSETVKIHELSPVR 355
Db 324 DQGLTLVNNPDSGATEEV-AGVNDNSMEIGFNAKYLDDITSQSLGEDA-IFLLADAG 381
QY 356 PFTL---TPGDEESFIQILTPVR 376
Db 382 SPTLVDRDTAGDDA---LYVLMFMR 402

RESULT 25
DNA polymerase III, beta chain (dnan) RP419 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 09-Jul-2004
C:Accession: B71700
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, U.
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499; PMID:9823893
A:Accession: B71700
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-381 <AND>
A:Cross-references: UNIPROT:Q9ZDB3; GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAA1487
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: dnan; RP419
C:Superfamily: DNA-directed DNA polymerase III beta chain

Query Match 17.3%; Score 323; DB 2; Length 381;
Best Local Similarity 24.6%; Pred. No. 8.2e-14;
Matches 96; Conservative 87; Mismatches 182; Indels 26; Gaps 10;

QY 1 MIQFSINRFLTHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNQISIENTIPV 60
Db 1 MLKLIVETKTLQSLGFARSITEKENVPEYANIKLSAQDGNLESLSTNMDLYLSQKIAV 60
QY 61 SNENAGLLITSPGAILLERASFFINISSLPDISINVKEIQHVLTGSKSITILKGKDV 120
Db 61 Q-----VLNGETVATQTLSDIVRFPDPSBELTLTEITQLEIKGQNCCKFNFLTLPV 113
QY 121 DOYPRLOEVSTENPLILKTKLLKSLIAETAFASLOESRPILTGVHIVLSNHKDFKAVAT 180
Db 114 SFPFAMDSIKPVSFKISCADPAKIESTKFSISLDETRYNLNGIYLIHKD-KEPFAAST 172
QY 181 DSHRMSQRLITLIDNTSADLMVLPSPKSLREFSAVFTDD---IETVEFFSPSOILF-RSE 236
Db 173 DGYRLISWITLLEKIKPGVILPQKSAEILKIVKDPKNIHEDIEILLSSNKIKFICNE 232
QY 237 HISFYTRLEGNYPDRLMTETETEVVFNQSLRHAMERAFILSNATQNGTVKLEITQ 296
Db 233 NTILSKLIDGTFDQYSAFIPKSSSVSKLVNKRKIFADSIERTAITI-VEKPRAVKLSLR 291
QY 297 NHI-----SAHVNSPVGKVNEDLDIVSQSGD---LTISFNPTVILIESLKAIKSETVKI 348
Db 292 KILEISAVGEAGTAKIEITASQDKESFYENHNDESIVGFNPQYLEDVLAIKASDIVEL 351
QY 349 HF--LSPVRPFTLT-PGDEESFIQILTPVR 376
Db 352 YFSDISASAPVLKIFRNPDKOIFV--IMPVK 380
```

```
RESULT 26
E72400
DNA polymerase III, beta subunit - Thermotoga maritima (strain MS98)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: E72400
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: E72400
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-366 <ARN>
A:Cross-references: UNIPROT:Q9WYA0; GB:AE001708; GB:AE000512; NID:g4980740; PIDN:AAD35355
A:Experimental source: strain MS98
C:Genetics:
A:Gene: TM0262
C:Superfamily: DNA-directed DNA polymerase III beta chain

Query Match 17.1%; Score 319.5; DB 2; Length 366;
Best Local Similarity 23.4%; Pred. No. 1.3e-13;
Matches 85; Conservative 97; Mismatches 167; Indels 15; Gaps 9;

QY 15 LNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNQISIENTIPVSNENAGLLITSPGA 74
Db 14 ITIASKALAKSKVKPILAGFLFEVKDGNFYCATDLETGVKATVNAAEISGEARFVWPGD 73
QY 75 ILLEASFFINISSLPDISINVKEIQHVLTGSKSITILKGKDVQYPRLOEVSTENP 134
Db 74 VIQK-----MVKVLPD-EITELSLEGDALVISSGSTVFRITTPADEPFEITPAESGIT 126
QY 135 LILKTKLLKSLIAETAFASLOESRPILTGVHIVLSNHKDFKAVATDSHRMSQRLITLD 193
Db 127 FEVDTSLLLEEMVEKVIFAAAKDEFMRNLNGVFWEL--HKNLRLVASDGFRLAAEQIE 184
QY 194 NTSADLMVLPSPKSLREFSAVFTDDIE-TVEFFSPSOILFRSEHISFYTRLEGNYPDT 252
Db 185 N-EEASFLSLSKMKEVQNVLNDTTEPTITVRYDGRVSLSTNDVETVMRVVDAEPDY 243
QY 253 DRLMTETETEVVFNQSLRHAMERAFILSNATQNGTVKLEITQNHISAHVNSPVGKYN 312
Db 244 KRVIPETPKTKVWSRKELRESLKRVMVVIASKSE-SVKFEIEENVMRLVSKSPDYGEVV 302
QY 313 EDLDIVSQSGDLTISFNPTVILIESLKAIKSETVKIHELSPVRPFTLTGDEESFIQIL 372
Db 303 DEVE-VQKEGEDLVIAFPKFIEDVLKHETEIEEMNFVDSTSPCOINPLD-ISGYLYIV 360
QY 373 TPVR 376
Db 361 MPIR 364

RESULT 27
DUECB
DNA-directed DNA polymerase (EC 2.7.7.7) III beta chain - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 28-Aug-1985 #sequence_revision 31-Mar-1988 #text_change 09-Jul-2004
C:Accession: A91510; A93996; A22168; I41190; I54000; F65172; A00719; B24944
R:Ohmori, H.; Kimura, M.; Nagata, T.; Sakakibara, Y.
Gene 28, 159-170, 1984
A:Title: Structural analysis of the dnaA and dnaN genes of Escherichia coli.
A:Reference number: A91510; MUID:84237568; PMID:6234204
A:Accession: A91510
A:Molecule type: DNA
A:Residues: 1-366 <OHM>
A:Cross-references: UNIPROT:P00583; GB:J01602; NID:g145758; PIDN:AAB59150.1; PID:g145761
A:Experimental source: strain K-12
R:Blancat, M.A.; Sandler, S.J.; Armengod, M.E.; Ream, L.W.; Clark, A.J.
Proc. Natl. Acad. Sci. U.S.A. 81, 4622-4626, 1984
```



A;Title: Molecular analysis of the recF gene of Escherichia coli.

A;Reference number: A93996; MUID:84272685; PMID:6379647

A;Accession: A93996

A;Molecule type: DNA

A;Residues: 297-366 <BLA>

A;Cross-references: GB:K02179; NID:g147537; PIDN:AAA24510.1; PID:g147538

A;Adachi, T.; Mizuuchi, K.; Menzel, R.; Gellert, M.

Nucleic Acids Res. 12, 6389-6395, 1984

A;Title: DNA sequence and transcription of the region upstream of the E. coli gyrB gene.

A;Reference number: A22168; MUID:84297235; PMID:6089112

A;Accession: A22168

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 342-366 <ADA>

A;Cross-references: GB:X04341; GB:X00870; NID:g41643; PIDN:CAA27869.1; PID:g41644

A;Experimental source: strain K-12

R;Armengod, M. 263, 12109-12114, 1988

A;Title: Transcriptional Organization of the dnaN and recF Genes of Escherichia coli K-12

A;Reference number: I41190; MUID:88298898; PMID:2841344

A;Accession: I41190

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-14 <RES>

A;Cross-references: GB:M19876; NID:g145779; PIDN:AAA23695.1; PID:g551800

R;Armengod, M.

Gene 43, 183-196, 1986

A;Title: Overlapping arrangement of the recF and dnaN operons of Escherichia coli; Position

A;Reference number: I54000; MUID:86301872; PMID:3527871

A;Accession: I54000

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 143-170 <RE2>

A;Cross-references: GB:M13822; NID:g147541; PIDN:AAA24512.1; PID:g147542

R;Blattner, F. R.; Plunkett III, G.; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Co

.A.; Rose, D. J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: F65172

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-366 <BLAT>

A;Cross-references: GB:AR000447; GB:U00096; NID:g2367266; PIDN:AAC76724.1; PID:g1790136;

A;Experimental source: strain K-12, substrain MG1655

C;Genetics:

A;Gene: dnaN

A;Map position: 83 min

A;Complex: DNA polymerase III is a multichain complex; alpha, epsilon, theta, tau, gamma

C;Function: <GEN>

A;Description: DNA synthesis; synthesizes both, the lagging and the leading strands in E

A;Pathway: DNA biosynthesis

A;Note: core enzymes (catalytic core) contains chains alpha, epsilon and theta; alpha cha

required for assembly; tau allows dimerization of the core complex and processivity is

te; the holoenzyme is completed by addition of beta chain which clamps the enzyme to DNA

C;Function: <BET>

A;Description: beta chain is required for initiation of replication

A;Note: can slide along duplex DNA bidirectionally and ATP-independent; binds core; cryst

C;Superfamily: DNA-directed DNA polymerase III beta chain

C;Keywords: DNA replication initiation; nucleotidyltransferase

Query Match 17.1%; Score 318; DB 1; Length 366;

Best Local Similarity 22.9%; Pred. No. 1.6e-13;

Matches 86; Conservative 93; Mismatches 186; Indels 10; Gaps 5;

QY 2 IQFSINRFLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNQISIENTIPVS 61

DB 1 MKFTVERHLLKPLQOVSGPLGGRTPLILGNLLQVADGTLTGTDLEMMVARVALV 60

QY 62 NENAGLLITSPGAILLEASFFINIISLPDISINVKIEHQHVLTSGKSEITLKGKDVD 121

DB 61 QPH-----EPGATVPARKFDICRGLPEGAIAVQLGERMLVRSGRSRLSTLPAA 114

QY 122 QYPRLOEVSTENPILITKYLKLSIIAETAFASLOESRPILTGVHIVLSNKHOKFVATD 181

DB 115 DFPNLDQSQSEVETPLPQATMKRLIEATQFSMAHQDVRYILNGM-LFETEGEELRTVATD 173

QY 182 SHRMSQRILITDNTSADLMVLPKSLREFSAVFTDDIETVEVFFSPSQILFRSEHISFY 241

DB 174 GHRLAVCSMPIGQSLPSHSHVIVPRKGVIELMRMLDGGDNPLRVQIGSNRIAHVGDFTFT 233

QY 242 TRLEGNYPTDRLMLTEFETEVVFNQSLRHHAMERAFILSNATQNGTVKLEITQNHISA 301

DB 234 SKLVDRFPDYRRVLPKPNPKHLEAGCDLLKQAFARAAILSNEKFRG-VRLVSVSENQLKI 292

QY 302 HVNSPEVQKVNEDLDIVSQGSDLTISFNPTYLIESLKAIKSETVKIHFLSPVPFTLTP 361

DB 293 TANNPEQEAEEILD-VTYSGAEMEIGFNVSIVLDVNLKCNVRMMLTDSVSSVQIED 351

QY 362 GDEESESFIQLITPVR 376

DB 352 A-ASQSAAYVVMVPMR 365

#### RESULT 28

D91208

DNA polymerase III beta-subunit [imported] - Escherichia coli (strain O157:H7, substrain

C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004

A;Accession: D91208

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C. G.

gasawara, N.; Yasunaga, T.; Kuahara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno

A;Reference number: A99629; MUID:21156231; PMID:11258796

A;Accession: D91208

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-366 <HAY>

A;Cross-references: UNIPROT:P00583; GB:BA000007; PIDN:BA838059.1; PID:g13364111; GSPDB:52

A;Experimental source: strain O157:H7, substrain RIMD 0509952

C;Genetics:

A;Gene: EC94636

C;Superfamily: DNA-directed DNA polymerase III beta chain

Query Match 17.1%; Score 318; DB 2; Length 366;

Best Local Similarity 22.9%; Pred. No. 1.6e-13;

Matches 86; Conservative 93; Mismatches 186; Indels 10; Gaps 5;

QY 2 IQFSINRFLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNQISIENTIPVS 61

DB 1 MKFTVERHLLKPLQOVSGPLGGRTPLILGNLLQVADGTLTGTDLEMMVARVALV 60

QY 62 NENAGLLITSPGAILLEASFFINIISLPDISINVKIEHQHVLTSGKSEITLKGKDVD 121

DB 61 QPH-----EPGATVPARKFDICRGLPEGAIAVQLGERMLVRSGRSRLSTLPAA 114

QY 122 QYPRLOEVSTENPILITKYLKLSIIAETAFASLOESRPILTGVHIVLSNKHOKFVATD 181

DB 115 DFPNLDQSQSEVETPLPQATMKRLIEATQFSMAHQDVRYILNGM-LFETEGEELRTVATD 173

QY 182 SHRMSQRILITDNTSADLMVLPKSLREFSAVFTDDIETVEVFFSPSQILFRSEHISFY 241

DB 174 GHRLAVCSMPIGQSLPSHSHVIVPRKGVIELMRMLDGGDNPLRVQIGSNRIAHVGDFTFT 233

QY 242 TRLEGNYPTDRLMLTEFETEVVFNQSLRHHAMERAFILSNATQNGTVKLEITQNHISA 301

DB 234 SKLVDRFPDYRRVLPKPNPKHLEAGCDLLKQAFARAAILSNEKFRG-VRLVSVSENQLKI 292

QY 302 HVNSPEVQKVNEDLDIVSQGSDLTISFNPTYLIESLKAIKSETVKIHFLSPVPFTLTP 361

DB 293 TANNPEQEAEEILD-VTYSGAEMEIGFNVSIVLDVNLKCNVRMMLTDSVSSVQIED 351

QY 362 GDEESESFIQLITPVR 376

DB 352 A-ASQSAAYVVMVPMR 365



## RESULT 29

AF6054  
DNA polymerase III, beta-subunit [imported] - Escherichia coli (strain O157:H7, substrain C)  
C:Species: Escherichia coli  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: F6054  
R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew, J.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: F6054  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-366 <STO>  
A:CROSS-references: UNIPROT:P00583; GB:AE005174; NID:G12518540; PIDN:AAG58898.1; GSPDB:G12518540  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: dnaN  
C:Superfamily: DNA-directed DNA polymerase III beta chain

Query Match 17.1%; Score 318; DB 2; Length 366;  
Best Local Similarity 22.9%; Pred. No. 1.6e-13;  
Matches 86; Conservative 93; Mismatches 186; Indels 10; Gaps 5;  
QY 2 IQFSINRTLFTHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNQISIENTIPVS 61  
DB 1 MKFTVEREHLKPLQOVSGPLGGRTPLPILGNLLQVADGTLSTLGTDLMEMVARVALV 60  
QY 62 NENAGLLITSPGAILLEASFFINISSLPDISINVKIEHQVLTSGKSEITLKGKVD 121  
DB 61 QPH-----EPGATVPARKFPDPCRGLEGAIEAVQLEGERMLVRSGRSFLSTLPAA 114  
QY 122 QYPRLOEVSTENPLILKTKLSIIAETAFASLOESRPILTVGHIVLSNKHDPKAVATD 181  
DB 115 DPNLDDWQSEVETPLQATMKRLTEATQFSMAHQDVRYLNGM-LFETEGELTVAID 173  
QY 182 SHRMSQRLITLNTSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSHISFY 241  
DB 174 GHRLAVCSMPIGQSLPSHVSIVPRKGVLMELMDGGDPLRVQIGSNIRAHVGDFT 233  
QY 242 TRLEGNPTDRLMTFETEVVNTQSLRHAMERAFILSNATONGTVKLEITONHISA 301  
DB 234 SKLVDRFPDYRRVLPKPNPKHLEAGCDLLKQAFARAAILSNKFRG-VRLYVSENOLKI 292  
QY 302 HNSPEVGKVNEDDIVSOGSDLTISFNPTVLTIESLKAISKSETVKIHFSLSPVPFTLTP 361  
DB 293 TANNPEQEAEEILD-VTYSGAEMEIGFNVSYLDVNLKCNVRRMLTDSVSSVQIED 351  
QY 362 GDEESFQILTPVR 376  
DB 352 A-ASQSAAYVVMR 365

## RESULT 30

AF0497  
DNA-directed DNA polymerase (BC 2.7.7.7) [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AF0497  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell, Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AF0497  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-366 <KUR>  
A:CROSS-references: UNIPROT:Q8Z908; GB:AL590842; PIDN:CAC93546.1; PID:G15981986; GSPDB:G15981986

## C:Genetics:

A:Gene: dnaN  
C:Superfamily: DNA-directed DNA polymerase III beta chain  
C:Keywords: nucleotidyltransferase

Query Match 17.0%; Score 317; DB 2; Length 366;  
Best Local Similarity 24.0%; Pred. No. 1.9e-13;  
Matches 91; Conservative 87; Mismatches 183; Indels 18; Gaps 6;  
QY 2 IQFSINRTLFTHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNQISIENTIPVS 61  
DB 1 MKFTIEREHLKPLQOVSSPLGGRTPLPILGNLLQVTEGSLRLTGTDLMEMVACVALS 60  
QY 62 NENAGLLITSPGAILLEASFFINISSLPDISINVKIEHQVLTSGKSEITLKGKVD 121  
DB 61 QSH-----EPGATVPARKFPDIWRGLPEGAETVALDGRLLVRSGRSFLSTLPAL 114  
QY 122 QYPRLOEVSTENPLILKTKLSIIAETAFASLOESRPILTVGHIVLSNKHDPKAVATD 181  
DB 115 DPNLDDWQSEVETPLQATMKRLTEATQFSMAHQDVRYLNGM-LFETEGELTVAID 173  
QY 182 SHRMSQRLITLNTSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSHISFY 241  
DB 174 GHRLAVCSMPIGQSLPSHVSIVPRKGVLMELMDGGDPLRVQIGSNIRAHVGDFT 233  
QY 242 TRLEGNPTDRLMTFETEVVNTQSLRHAMERAFILSNATONGTVKLEITONHISA 301  
DB 234 SKLVDRFPDYRRVLPKPNPKHLEAGCDLLKQAFARAAILSNKFRG-VRLYVSENOLKI 292  
QY 302 HNSPEVGKVNEDDIVSOGSDLTISFNPTVLTIESLKAISKSETVKIHFSLSPVPFTLTP 361  
DB 293 TANNPEQEAEEILD-VSYEGTEMEIGFNVSYLDVNLKCNVRRMLTDSVSSVQI-- 349  
QY 362 GDEESFQI-----LITPVR 376  
DB 350 --EDSASQAAAYVVMR 365

## RESULT 31

AF2535  
DNA polymerase III beta chain [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C:Accession: AF2535  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.; DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AF2535  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-456 <KUR>  
A:CROSS-references: UNIPROT:Q8ZSE2; GB:AP003602; PIDN:BAB77212.1; PID:G17134654; GSPDB:G17134654  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: alr7569  
A:Genome: plasmid

Query Match 16.9%; Score 315.5; DB 2; Length 456;  
Best Local Similarity 24.9%; Pred. No. 3.3e-13;  
Matches 96; Conservative 85; Mismatches 179; Indels 25; Gaps 12;  
QY 2 IQFSINRTLFTHALNTTKRAISTKNAIPILSSIKIEVTSTG-VTITGSGNQISIENTIPV 60  
DB 1 MKLSIEQSLAEILETAYFALSPKPTDPLGNILLIANEDGIVSATGNLNTIHTTTA 60  
QY 61 SNENAGLLITSPGAILLEASFFINISSL-PDISINVKIEHQVLTSGKSEITLKGKD 119  
DB 61 NVETS-----GOVALPAKLLTDTINNVRGEITL---EVENQACIITHNSGKRLIGK 110  
QY 120 VDOQPRLOEVSTENPL--ILTKLLKLSIIAETAFASLOESRPILTVGHIVLSNKHDPK 177

Db 111 PDEFTLPK--GENPIEVLNLSAKLQAEGTLCTNGDETCLVLTGVNFKIDTNK-WQA 167  
QY 178 VATDSHRMSQRLTLTDLNTSAD-LMVVLPSKSLRERSAVFTDDIET--VEVFFSPSOLIPR 234  
Db 168 ASINGHKLALVCTLSEVSDPFDFTVPKSLSELSKILSQSADTVSCVNLISNKTIEFS 227  
QY 235 SEHISFYRLLEGNYPDTRLMLTEFETEVVNTQSLRHAMERAFILISNATNGTVKL-- 292  
Db 228 LPHTKVISRLLEGYPKINSILIPRTFEYFTLKERGFESALKRVSLAERKQK-VVKILW 286  
QY 293 EITQNHISAHVNSPEVKVNE--LDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHF 350  
Db 287 ELEATOATLYTEATDGDVADSVLMKPAIHNSENISIGLNIDYLLGKLKHISTDIIVRC 346  
QY 351 LSPVRPFTLTPGDEESFIQITPV 375  
Db 347 NKPTQPVICPMGGLNLQYLWNPV 371

RESULT 32  
AD2057  
DNA polymerase III beta chain [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C:Accession: AD2057  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AD2057  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-387 <KUR>  
A:Cross-references: UNIPROT:Q8YVCG; GB:BA000019; PIDN:BA073709.1; PID:gl7131100; GSPDB:G  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: dnaN  
C:Superfamily: DNA-directed DNA polymerase III beta chain

Query Match 16.9%; Score 314; DB 2; Length 387;  
Best Local Similarity 25.1%; Pred. No. 3.3e-13;  
Matches 92; Conservative 79; Mismatches 166; Indels 30; Gaps 10;

QY 15 LMTTKRAISTKNAIPILSSI--KIEVTGTGVTLTGSGQISIENTIPVSNENAGLITSP 72  
Db 14 LSLVNRAVPSRTHFVLNVLQDAETNQVSLTAFDLSIGIRISF-----NAD--VWQS 66  
QY 73 GAILEASFFNITSSLPDISINVKI-----EQHVLTSGKSEITLKGKVDQY 123  
Db 67 GAIALPAKLLVDITSRLPEGETLDDDESATDGTATGEGLIVSLTPKTKYQLRAMGAGEF 126  
QY 124 PRLOQVSTENPLILTKLKSIIAETAFASLQSRPILTGTHVILSNHDKFKAVATDSH 183  
Db 127 PELPLIENTAIYLTATSLIGRLSPATSGDETQKVLGTGHLTV-KQDTLEFAATDGH 185  
QY 184 RMSQRLIT----LNTSADLMVLPSKSLRERSAVFTDDI---ETVEVFFSPSOLIPRSE 236  
Db 186 RLAVVETTNERPLEDNDQOEVTVPARALRELERMLAHNAASEPIALYDQGOVFAWQ 245  
QY 237 HISFYRLLEGNYPDTRLMLTEFETEVVNTQSLRHAMERAFILISNATNGTVKLHI-- 294  
Db 246 NORLTSRTLEGYQAYROLIQRQERQVITERRQFLSTLERIAVLAD-QKNNIVKLITDS 304  
QY 295 TQNHISAHVNSPEVKVNE--LDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFSPV 354  
Db 305 TAQELTSLCEAQWGSGRSNV-AEIAGEDIEIAFNVKYLMLEGLKALPSPFIQNHINQL 363  
QY 355 RPFTLTP 361  
Db 364 TPVIFTTP 370

## RESULT 33

AH3494

DNA-directed DNA polymerase (EC 2.7.7.7) [imported] - Brucella melitensis (strain 16M)  
C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 09-Jul-2004

C:Accession: AH3494

R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova, M.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis

A:Reference number: AD3252; PMID:11756688

A:Accession: AH3494

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-397 <KUR>

A:Cross-references: UNIPROT:Q8YED6; UNIPROT:Q8G3E6; GB:AE008917; PIDN:AALS3123.1; PID:gl

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BMRI1942

A:Map position: I

C:Superfamily: DNA-directed DNA polymerase III beta chain

C:Keywords: nucleotidyltransferase

Query Match 16.9%; Score 314; DB 2; Length 397;  
Best Local Similarity 24.1%; Pred. No. 3.4e-13;  
Matches 93; Conservative 89; Mismatches 178; Indels 26; Gaps 11;

QY 2 IQFSINRFLTHALNTTKRAISTKNAIPILSSIKIEVTGTGVTLTGSGQISIENTIPVS 61  
Db 26 MRVTLERSNLLKSLNHVRVVERENTPIILSNVLQAEAGASLAKMKTDLDEL----- 78  
QY 62 NENAGLITSPGAILLEASFFNIISLPD-----ISINVKIEIHOHVLTTSKSEITLK 116  
Db 79 NEATAAMVEOAGATTVEPAHLLYDIVRKLPGAEVMLSTNP---DGGSMVISGKSPRLQ 135  
QY 117 GKVDVQYPRLOQVSTENPLILTKLKSIIAETAFASLQESRPILTV--HIVLSNHK- 173  
Db 136 CLPQSDPELTAGATPHTSFRIEAAQLKLLIDRTQFAISTEETRYLNGIFHAIESGAL 195  
QY 174 DFKAATDSHRMSQ-RLITLDTNDSADLMVLPSKSLRERSAV--FTDDIETVEVFFSPQ 230  
Db 196 KLRATATDGHRLARAEAPSGTEGMPGIIIPKTVAELOKLVDPVDDVVTVL--SDAK 253  
QY 231 ILRSEHISFYRLLEGNYPDTRLMLTEFETEVVNTQSLRHAMERAFILISNATNGTV 290  
Db 254 IRTVGSVLTSLKIDGTFPDYQYRVPISGNDKKLTIDRQDFAASVDRVSTIS--SERGRAV 312  
QY 291 KLEITQNHISAHVNSPEVKVNE--LDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHF 350  
Db 313 KLSIADGQLTLTVNPPDSSGATDEL-AADYDGPDLIGNSKYLDD-ITGQLSGTDAVFM 370  
QY 351 LSPVRPFTLTPGDEESFIQITPV 376  
Db 371 LADAGSPTLVRDGTGDBDVLVYLMFMR 396

## RESULT 34

JQ0734

DNA-directed DNA polymerase (EC 2.7.7.7) III beta chain - Proteus mirabilis  
C:Species: Proteus mirabilis

C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 09-Jul-2004

C:Accession: JQ0734

R:Skovgaard, O.

Gene 93, 27-34, 1990

A:Title: Nucleotide sequence of a Proteus mirabilis DNA fragment homologous to the 60K-r  
A:Reference number: JQ0729; MUID:91033012; PMID:2172087

A:Accession: JQ0734

A:Molecule type: DNA

A:Residues: 1-367 <SKO>

A:Cross-references: UNIPROT:P22838; GB:M58352; GB:M31295; NID:g150873; PIDN:AAA03959.1;

A:Experimental source: strain LM1509

```

Db      173 GKRLAKIDTWISLDPFSGD--YIIPKAVEIIRPASEDVQST-FLDQTKIAVECNT 229
Qy      239 SFYTRLLEGNYPTDRLMLMTFETFEVVFNTQSLRHAMERAPLISNATQNGTVKLEITQNH 298
Db      230 LLVTKLLSGSEFPDFSPVISTQSSVQLNLHREELISLLKQVAFETNESSH-SVKRFSFSGE 288
Qy      299 ISAHVNSPEYGVKNEDLDIVSQSGDLTISFNPTLYLESLSKAIKSETVKIHFSLSPVPFPT 358
Db      289 LTLTANCTKVGEGKVSMA-VNYTGETLEIAFNPFPLDLKHSRDELVLQGISDSYNPGI 347
Qy      359 LTPGDEEESFIQITPVR 376
Db      348 IT--DSTRSLF-VIMPMR 362

RESULT 36
B41870
C:Directed DNA polymerase (EC 2.7.7.7) III beta chain - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 09-Jul-2004
C:Accession: B41870; T10965
R:Calcutt, M.J.; Schmidt, F.J.
J. Bacteriol. 174, 3220-3226, 1992
A:Title: Conserved gene arrangement in the origin region of the Streptomyces coelicolor
A:Reference number: A41870; MUID:92250416; PMID:1577691
A:Accession: B41870
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <CAL>
A:Cross-references: UNIPROT:P27903; GB:M82836; NID:G6539740; PIDN:AAA26735.1; PID:gi15322
Gene 151, 23-28, 1994

```

A;Reference number: 217435; MOLD:95129885; FMD:7828880  
 A;Accession: T10965  
 A;Status: Preliminary; translated from GE/EMBL/DBD  
 A;Molecule type: DNA  
 A;Residues: 366-376 <CA2>  
 A;Cross-references: EMBL:L27063; NID:g436023; PID:g777777  
 A;Experimental source: strain A3(2)  
 C;Genetics:  
 A;Gene: dhaN  
 A;Start codon: GTG  
 C;Superfamily: DNA-directed DNA polymerase III beta chain  
 C;Keywords: nucleotidyltransferase  
  
 Query Match 15.9%; Score 296; DB 2; Length 376;  
 Best Local Similarity 24.6%; Pred. No. 4.7e-12;  
 Matches 96; Conservative Indels 34; Gaps 11;  
  
 OV 2 I0FSINRTLFIHALNTTKRAISTKNAIPLISSIKIEVTSTGVLTGNSGOISI-----EN 56

```

1 MKIRVERDVLTAEAVAMAAARSLPAPPPAVL-----GLLKABEGQLUSSSFYEV 51
QY 57 TIPVSNENAGLLITSPCAILLAEAFFNISSIPDISINVKEIEHQHVLVLTSGKSEITLK 116
Db 52 SARVSVEAE---IEEGTVLVSGRLLADISRALPNRPVEI-STDGVRAVTVCGSSRFTLH 107
QY 117 GKVDVQYPRLOEVSVENTENPLIKTKLKSIIAETAFAASLOESRPILTCGHVILSNHHDKF 176
Db 108 TLPVEEYPALPQM-PEATGTVPGVEFASAVQQAAGRODTLPVLTGVRIEBG-DSVT 165
QY 177 AVATDSHRMSQR--LITLDNTSADLMVVLPSKSLREFSAVFTDDIETVEVFFFSQ---- 230
Db 166 LASTDTRYFAVRBFLWKPNPDITSAAVALVPKTLQDTAKALTSQDQVILALSGSGAGEGL 225
QY 231 ILFRSEHISFYTLLEGNYPDTRLMLTBEFETEVFNFTQSLRHAMERAFILSNATQGTV 290
Db 226 IGFEGAGRRITTRLLEGDLPKYKTLPTFTEFNSAVVIETAPFVEAVKRVLV--AERTPV 283
QY 291 KLSITQNHISAHVNSPEVGKVNEDDLIVSOGSGDLTISFNPTYLIESKAIKSTVTKHF 350
Db 284 RLSFEOCVLILEAGSSDDAQAVSRVD--AQLEGDDISIAFNPTFLDGLGIDSAPVQSLF 342

```

Qy	351 LSPVPFTLT-----PGDEEESFIQLITPVR 376
	: :   } : : : :
Db	343 TTSTKPALLSGRPDAVDAEAAYKYLIMPVR 373

RESULT 37

H72090  
DNA polymerase III (beta chain) - Chlamydomydia pneumoniae (strain CWL029)  
C:Species: Chlamydomydia pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 09-Jul-2004  
C:Accession: H72090  
R:Kalan, S.; Mitchell, W.; Marxthe, R.; Lammell, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606; PMID:10192388  
A:Accession: H72090  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-366 <ARN>  
A:Cross-references: UNIPROT:Q9Z8K0; GB:AE001618; GB:AE001363; NID:94376610; PIDN:AAD1848  
A:Experimental source: strain CWL029  
C:Genetics:  
A:Gene: dnaN  
C:Superfamily: DNA-directed DNA polymerase III beta chain

Query Match	15.9%;	Score 295.5;	DB 2;	Length 366;
Best Local Similarity	22.5%;	Pred. No. 4.9e-12;		
Matches 85; Conservative	89;	Mismatches 186;	Indels 17;	Gaps 0

QY	2	I QFSINRFLFHALNTTKRAISTKNAIPILSSIKIEVTSGVTLTCSGNOQI3IENTIPWS	61
Db	1	MFVSVRNELGNLKKIQSVVPQNPPIVLTHVLITYNDELVFATDILTVCSTCVAK	60
QY	62	NENAGLLITSPGAILLEASFINIISLPPDISINVKIEBOHQVVTLSKSEITLKGDOVD	121
Db	61	-----VYEGKALISPKRFFOLVKELTANLEISSAGEMAQITSGSCFRLLSMEKE	113
QY	122	OYPRLOEVSTENPILKTLKLSIAETAFAASLQESRPILTGVHVLVSNHKDFKAVATD	181
Db	114	DFPWLPIQONALRSLPAEQKTMWLQRTSPAVSREESRYVLTVGLVLATANGV-ATIVGTD	172
QY	182	SHRMSQ--RUITLNTSADLMVVLPSKLSREFSAVFTDDIETVEVFFSPQILFRSEHIS	239
Db	173	GKRLAKDAEVLTDKPSG-EYIIPKAVEEIIKMSDSEAA--FLOODKIAVECDNTL	230
QY	240	FYTELLSGNPDDTRLMLTTEFEVVFNTQSRHAMERAPFLISNATQNTVKLEITQNHII	299
Db	231	LITKLLSGEPDFSPVISTESNVKGLDHRBELITLLKQVALFTNESSH-SVKFSPFLPCEL	289
QY	300	SAHNSPEVGVKNEDDIDVSGSDSLTISFNPTYLIESLKATKSTVKIHLSPVRPFTL	359
Db	290	TLTANCTKVGEKVSM-AVNSGELLEIAFNFFFLDLILKSKDELVSIGISDSYNPGII	348
QY	360	TPGDEESFTQIITPVR	376
Db	349	T--DSASGLFVIMPMR	362

RESULT 38

K850311 36  
 B85533  
 DNA polymerase III (beta chain) [imported] - Chlamydomonadales (strain J138)  
 C:Species: Chlamydomonadales; Chlamydomonadales  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
 C:Accession: B85533  
 R;Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Ouchi, K.; Shiba, T.; Ise  
 Nucleic Acids Res. 28, 2311-2314, 2000  
 A:Title: Comparison of whole genome sequences of Chlamydomonadales J138.  
 A:Reference number: A86491, PMID:20330349; PMID:10871362  
 A:Accession: B85533  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-366 <STO>

A;Cross-references: UNIPROT:Q29XK0; GB:BA000008; NID:g8978712; PIDN:EAA98548.1; GSPD8:GN  
A;Experimental source: strain J138  
C;Genetics:  
A;Gene: dhaN  
C;Superfamily: DNA-directed DNA polymerase III beta chain

Query Match	15.9%;	Score 295.5;	DB 2;	Length 366;
Best Local Similarity	22.5%;	Pred. No. 4.9e-12;		
Matches 85; Conservative	89;	Mismatches 186;	Indels 17;	Gaps 8;

QY	2	I QFSINRFLTHALNTTKRAISTKNAIPILSSIKIEVTSTGVTILTSGNQISIENTIPVS	61
Db	1	MKFVVSRNELGNLIKKIQSVVPQNTPVLTTHVLITETYNDELVFTATDLTVSTRCVTKAK	60
QY	62	NENAGLLITSPGAILLEASPINIISLPDISINVKIEBOHQVVLITSGKSEITUKGKQVD	121
Db	61	-----VYEGKAIISIPSKRFFQVVKELTANLEISSAGEMAQITSGSCFRLLSMEKE	113
QY	122	QYPRLOQVSTENPIILTKLLKIIAETAFAASLOESRPILTGVHIVLSNHKDFKAVATD	181
Db	114	DFPWLPIQNALRFSLPAEQKLTMLQRTSFVARSERYVLTGVLLAIANG-ATVIGTD	172
QY	182	SHRMSQ--RLITLNTSGADLMVLPKSLREFSAVFTDDIETVEVFPSSQILFRSEHIS	239
Db	173	GKRLADAEVLTDLKPSFG-EYIIPIKAVEEIIKMCSEGEAT-IFLDQKIAVECDNTL	230
QY	240	FYTELLEGNVPDTRLLMTTFEVEVVENTOSLRHAMERAFILSNATONGTVKLEITQNH	299
Db	231	LITKLLSGEPDFSPVISTESNVKLDLHREELITLTKQVALFTNESSH-SVKFSFLPGEL	289
QY	300	SAHVSPEVGKWNEDLDIVSQSGDLTISNPVYLIESLKAISKETVKIHLFSVPVFTL	359
Db	290	TLTANCTKVGEKVSM-ANVYSGELLEIAENPFFDLILKSKDELVSIGISDSYNPGII	348
QY	360	TPGDEEESFTQITTPVR	376
Db	349	T---DSASGLFVIMPMR	362

RESULT 39

F81578  
DNA polymerase III, beta chain CP0419 [imported] - Chlamydothrix pneumoniae (strain AR39  
C:Species: Chlamydothrix pneumoniae, Chlamydothrix pneumoniae  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
C:Accession: F81578  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydothrix trachomatis MoPn and Chlamydothrix pneumoniae AR39.  
A:Reference number: A81500; MUID:20150255; PMID:10684935  
A:Accession: F81578  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-366 <REA>  
A:Cross-references: UNIPROT:Q928K0; GB:AF002203; GB:AE002161; NID:g7189339; PIDN:AAF3826  
A:Experimental source: strain AR39, HL cells  
C:Genetics:  
A:Gene: CP0419  
C:Superfamily: DNA-directed DNA polymerase III beta chain

Query Match 15.8%; Score 293.5; DB 2; Length 366;  
Best Local Similarity 22.5%; Pred. No. 6.6e-12;  
Matches 85; Conservative 89; Mismatches 186; Indels 17; Gaps 8;

```

Qy 2 IQFSINTLFIHALNTTKRAISTKNAIPILSSKIEVTSSTGVLTGSGQISIENTIPVS 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MKFVSRNELGNLIKQISWPQNTPIPVLTHTLJETYNDELVFATDILTSTRCVTKAK 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 62 NENAGLLITPGAILLEASFNITLSSLPDISINVKIEIQHCVLTSGKSEITLKGQVD 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 -----VYKGAISIPSRFFQVKETEANLEISYAGEMAQITSGSSCFRLLSMEKE 113
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 122 QYPRQVSTENPLIKTLKLSIIAETAFASLOESRPLTGVTHIVLSNHHKDFKAVATD 181

```

Search completed: January 28, 2005, 17:00:21  
Job time : 44 secs

```
Db 114 DFPMLPDQNALRSLPAEQKLTMLQRTSFVARSREESRYVLTGVLAAANGV-ATVGTGTD 172
Qy 182 SHRMSQ--RLITLONTSDADLMVLPSKSLRFSFAVFTDDIETVEVFFSPSOILFRSEHIS 239
Db 173 GKRLAKIDAETVLDKSPSG--EYIPIKAVEEIIKWCSEGEAT-IFLDQDKIAVECDNTL 230
Qy 240 FYTRLLEGNYPTDRLLMTETETEVNTQSLRHAMERAFILSNATONGTVKLEITQNH 299
Db 231 LITKLLSGEFPDFSPVISTESNVKLDLHREELITLLKQVALFTNESSH-SVKFSFLPGEL 289
Qy 300 SAHVNSPEVGVKNEDLDIVSQSGDLTISFNPTVLTIESLKAISKETVKIHFSLSPVRPFTL 359
Db 290 TLTANTCTKVGKGKVM-AVNTSGELLEIAFNPFPLDLKHSDKDELVSGLGSDSYNPGII 348
Qy 360 TPGDEESFIQLITPVR 376
Db 349 T--DSASGLFVIMPMR 362

RESULT 40
E71559
probable DNA pol III (beta chain) - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 09-Jul-2004
C:Accession: E71559
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac
A:Reference number: A71570; MUID:99000809; PMID:9784136
A:Accession: E71559
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-416 <ARN>
A:Cross-references: UNIPROT:O84078; GB:AE001282; GB:AE001273; NID:G3328466; PIDN:AAC6766
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
C:Gene: dnaN
C:Superfamily: DNA-directed DNA polymerase III beta chain

Query Match 15.7%; Score 292.5; DB 2; Length 416;
Best Local Similarity 22.7%; Pred No. 9.2e-12;
Matches 86; Conservative 91; Mismatches 183; Indels 19; Gaps 10;

Qy 1 MIQFSINRTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGTGTLTNGSQISIENTIPV 60
Db 50 IMKFVISRNELGNLIKQNVVPPQSTPIPLVTLVHIESCNDELVTATDLTVSTRCVKA 109
Qy 61 SNEAGLLITSPGAILLEASFFINIISLPDISINVKIEIOHQVVLTSCKSEITLKGDV 120
Db 110 K-----VYESGSVTIPSRFRFQILRELTAEIAHVAHSGEMATITSGSCFRLLSMGK 162
Qy 121 DOYPRLOEVSTENPLILATKLLKSTIAETAAASLOESRPILITGVHIVLSNHKDFKAVAT 180
Db 163 EDFPMLPDQNALRFTLDSERLKNFQRTSFVARSREESRYVLTGVLAAANGV-ATVGTGTD 221
Qy 181 DSHRMSQ--RLITLD-NTSADLMVLPSKSLRFSFAVFTDDIETVEVFFSPSOILFRSEH 237
Db 222 DGKRLAKIDTISLDPSPSGD--YIPIKAVEEIIKWCSEGEAT-IFLDQDKIAVECDNTL 278
Qy 238 ISFYTRLLEGNYPTDRLLMTETETEVNTQSLRHAMERAFILSNATONGTVKLEITQNH 297
Db 279 TLLVTKLLSGEFPDFSPVISTESNVKLDLHREELITLLKQVALFTNESSH-SVKFSFLPGEL 337
Qy 298 HISAVNSPEVGVKNEDLDIVSQSGDLTISFNPTVLTIESLKAISKETVKIHFSLSPVRPFTL 357
Db 338 ETLTANTCTKVGKGKVM-AVNTSGELLEIAFNPFPLDLKHSDKDELVSGLGSDSYNPGII 396
Qy 358 TLTGDEESFIQLITPVR 376
Db 397 IIT--DSTESLFLVIMPMR 412
```



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 16:59:44 ; Search time 146 Seconds  
(without alignments)  
935.393 Million cell updates/sec

Title: US-10-048-071-28  
Perfect score: 1863  
Sequence: 1 MIQSFNRTLFIHALNTTKR.....LTPGDEERSFIQLTPVPTN 378

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1608061 seqs, 361289386 residues

Total number of hits satisfying chosen parameters: 1608061

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1849	99.2	378	15 US-10-282-122A-74301	Sequence 74301, A
2	1527	82.0	378	15 US-10-282-122A-72018	Sequence 72018, A
3	1384	74.3	378	9 US-09-815-242-13446	Sequence 13446, A
4	1384	74.3	378	15 US-10-282-122A-73682	Sequence 73682, A
5	864.5	46.4	376	9 US-09-815-242-10907	Sequence 10907, A
6	864.5	46.4	376	15 US-10-282-122A-56984	Sequence 56984, A
7	862.5	46.3	376	15 US-10-282-122A-58007	Sequence 58007, A
8	748.5	40.2	377	15 US-10-282-122A-70833	Sequence 70833, A
9	735.5	39.5	377	15 US-10-282-122A-71908	Sequence 71908, A
10	730.5	39.2	377	9 US-09-815-242-12700	Sequence 12700, A
11	730.5	39.2	377	9 US-09-815-242-12726	Sequence 12726, A
12	730.5	39.2	377	14 US-10-282-287-10	Sequence 10, Appl
13	730.5	39.2	377	15 US-10-282-122A-44084	Sequence 44084, A

## ALIGNMENTS

### RESULT 1

US-10-282-122A-74301  
; Sequence 74301, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282.122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931

14	730.5	39.2	377	16	US-10-449-830A-2	Sequence 2, Appli
15	725.5	38.9	376	9	US-09-815-242-5669	Sequence 5669, Ap
16	703	37.7	381	15	US-10-282-122A-60333	Sequence 60333, A
17	692	37.1	379	15	US-10-282-122A-45628	Sequence 45628, A
18	586	31.5	334	15	US-10-671-403-174	Sequence 174, App
19	586	31.5	334	15	US-10-671-419-174	Sequence 174, App
20	586	31.5	334	15	US-10-670-844-174	Sequence 174, App
21	586	31.5	334	15	US-10-671-134-174	Sequence 174, App
22	586	31.5	334	15	US-10-672-098-174	Sequence 174, App
23	586	31.5	334	15	US-10-672-638-174	Sequence 174, App
24	586	31.5	334	15	US-10-673-127-174	Sequence 174, App
25	586	31.5	334	15	US-10-670-817-174	Sequence 174, App
26	586	31.5	334	15	US-10-673-119-174	Sequence 174, App
27	586	31.5	334	17	US-10-671-207-174	Sequence 174, App
28	457.5	24.6	366	15	US-10-282-122A-52020	Sequence 52020, A
29	448	24.0	402	15	US-10-282-122A-53348	Sequence 53348, A
30	442	23.7	375	15	US-10-282-122A-52377	Sequence 52377, A
31	380.5	20.4	399	15	US-10-282-122A-61654	Sequence 61654, A
32	374	20.1	387	15	US-10-282-122A-45981	Sequence 45981, A
33	364	19.5	402	9	US-09-712-363-147	Sequence 147, App
34	364	19.5	402	15	US-10-282-122A-64326	Sequence 64326, A
35	364	19.5	402	17	US-10-476-597-143	Sequence 143, App
36	363	19.5	402	15	US-10-282-122A-62773	Sequence 62773, A
37	362.5	19.5	367	15	US-10-282-122A-68144	Sequence 68144, A
38	362.5	19.5	367	15	US-10-671-403-112	Sequence 112, App
39	362.5	19.5	367	15	US-10-671-419-112	Sequence 112, App
40	362.5	19.5	367	15	US-10-670-844-112	Sequence 112, App
41	362.5	19.5	367	15	US-10-671-134-112	Sequence 112, App
42	362.5	19.5	367	15	US-10-673-098-112	Sequence 112, App
43	362.5	19.5	367	15	US-10-672-638-112	Sequence 112, App
44	362.5	19.5	367	15	US-10-673-127-112	Sequence 112, App
45	362.5	19.5	367	16	US-10-670-817-112	Sequence 112, App





```
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 74301
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-282-122A-74301

Query Match          99.2%; Score 1849; DB 15; Length 378;
Best Local Similarity 99.2%; Pred. No. 1.9e-144;
Matches 375; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MIQFSINRTLFHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNGQISIENTIPV 60
DB 1 MIQFSINRTLFHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNGQISIENTIPV 60

QY 61 SNEGALLITSPGAILLEASFFINIISLPDISINVKIEHOHVLTSGKSEITLKGKDV 120
DB 61 SNEGALLITSPGAILLEASFFINIISLPDISINVKIEHOHVLTSGKSEITLKGKDV 120

QY 121 DQYPRLOEVSTENPLILTKLLKSIIAETAFASLQESRPILTGVIHVLNHNKDFKAVAT 180
DB 121 DQYPRLOEVSTENPLILTKLLKSIIAETAFASLQESRPILTGVIHVLNHNKDFKAVAT 180

QY 181 DSHRMSORLITLNTSADLMVLPKSLREPSAVTDDIETVEVFPSPQILFRSEHISF 240
DB 181 DSHRMSORLITLNTSADLMVLPKSLREPSAVTDDIETVEVFPSPQILFRSEHISF 240

QY 241 YTRLLEGNYPTDRLMLTETEFETEVVNTQSLRHAMERAFILSNATONGVKLEITQNHIS 300
DB 241 YTRLLEGNYPTDRLMLTETEFETEVVNTQSLRHAMERAFILSNATONGVKLEITQNHIS 300

QY 301 AHVNSPEVGKVNEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFSPVRPFTLT 360
DB 301 AHVNSPEVGKVNEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFSPVRPFTLT 360

QY 361 PGDEESFIQLITPVRTN 378
DB 361 PGDEESFIQLITPVRTN 378

RESULT 2
US-10-282-122A-72018
; Sequence 72018, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
```

```
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 72018
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-282-122A-72018

Query Match          82.0%; Score 1527; DB 15; Length 378;
Best Local Similarity 79.9%; Pred. No. 8.5e-118;
Matches 302; Conservative 39; Mismatches 37; Indels 0; Gaps 0;

QY 1 MIQFSINRTLFHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNGQISIENTIPV 60
DB 1 MIQFSINRTLFHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNGQISIENTIPV 60

QY 61 SNEGALLITSPGAILLEASFFINIISLPDISINVKIEHOHVLTSGKSEITLKGKDV 120
DB 61 SNEGALLITSPGAILLEASFFINIISLPDISINVKIEHOHVLTSGKSEITLKGKDV 120

QY 121 DQYPRLOEVSTENPLILTKLLKSIIAETAFASLQESRPILTGVIHVLNHNKDFKAVAT 180
DB 121 DQYPRLOEVSTENPLILTKLLKSIIAETAFASLQESRPILTGVIHVLNHNKDFKAVAT 180

QY 181 DSHRMSORLITLNTSADLMVLPKSLREPSAVTDDIETVEVFPSPQILFRSEHISF 240
DB 181 DSHRMSORLITLNTSADLMVLPKSLREPSAVTDDIETVEVFPSPQILFRSEHISF 240

QY 241 YTRLLEGNYPTDRLMLTETEFETEVVNTQSLRHAMERAFILSNATONGVKLEITQNHIS 300
DB 241 YTRLLEGNYPTDRLMLTETEFETEVVNTQSLRHAMERAFILSNATONGVKLEITQNHIS 300

QY 301 AHVNSPEVGKVNEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFSPVRPFTLT 360
DB 301 AHVNSPEVGKVNEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFSPVRPFTLT 360

QY 361 PGDEESFIQLITPVRTN 378
DB 361 PGDEESFIQLITPVRTN 378

RESULT 3
US-09-815-242-13446
; Sequence 13446, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; PROKARYOTES
; FILE REFERENCE: ELITRA.011A
```

APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.0344  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 73682  
LENGTH: 378  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-10-282-122A-73682

Query Match 74.3%; Score 1384; DB 9; Length 378;  
Best Local Similarity 72.2%; Pred. No. 5.8e-106;  
Matches 273; Conservative 49; Mismatches 56; Indels 0; Gaps 0;  
QY 1 MIQFSINRFLTHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTSGNSQISIENTIPV 60  
DB 1 MHFSINKNLFQALNITKRAISSKNAIPILSTVKIDVTNEGVTLTSGNSQISIENTIPV 60  
QY 61 SNENAGLLITSPGAILLEASFFINISSLPDISINVKEIQHGVLTSGSKSEITLKGKDV 120  
DB 61 KNEAGLLITSLGSLLEASFFINIVSSLPDVTLDPEKEIQHGVLTSGSKSEITLKGKDV 120  
QY 121 DOYPRLOEVSTENPLILTKLLKSIIAETAPAASTQESRPILTVGHVLSNHKDPKAVAT 180  
DB 121 EOYPRIOISASTPILLETLLKLLKIIINETAFAASTQESRPILTVGHVLSNHKDPKAVAT 180  
QY 181 DSHRMSQRLITLIDNTSADLMVVLPSKSLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240  
DB 181 DSHRLSQKLLTEKNSDDFDVVPISRLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240  
QY 241 YTRLLEGNYPDTDRLLMTTEFTEVFNQSLRHAMERAFISNATQNGTVKLEITQNHIS 300  
DB 241 YTRLLEGNYPDTDRLLMTTEFTEVFNQSLRHAMERAFISNATQNGTVKLEITQNHIS 300  
QY 301 AHVNSPEVGVKNEDDIDVSGSGLTISFNPTLYLESKAIKSETVKHFLSPVPRFTLT 360  
DB 301 AHVNSPEVGVKNEDDIDVSGSGLTISFNPTLYLESKAIKSETVKHFLSPVPRFTLT 360  
QY 361 PGDEEESFIQLITPVRTN 378  
DB 361 PADTDEDFMQLITPVRTN 378

## RESULT 4

US-10-282-122A-73682  
Sequence 73682, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Lianguo  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.

APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.0344  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 73682  
LENGTH: 378  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-10-282-122A-73682

Query Match 74.3%; Score 1384; DB 15; Length 378;  
Best Local Similarity 72.2%; Pred. No. 5.8e-106;  
Matches 273; Conservative 49; Mismatches 56; Indels 0; Gaps 0;  
QY 1 MIQFSINRFLTHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTSGNSQISIENTIPV 60  
DB 1 MHFSINKNLFQALNITKRAISSKNAIPILSTVKIDVTNEGVTLTSGNSQISIENTIPV 60  
QY 61 SNENAGLLITSPGAILLEASFFINISSLPDISINVKEIQHGVLTSGSKSEITLKGKDV 120  
DB 61 KNEAGLLITSLGSLLEASFFINIVSSLPDVTLDPEKEIQHGVLTSGSKSEITLKGKDV 120  
QY 121 DOYPRLOEVSTENPLILTKLLKSIIAETAPAASTQESRPILTVGHVLSNHKDPKAVAT 180  
DB 121 EOYPRIOISASTPILLETLLKLLKIIINETAFAASTQESRPILTVGHVLSNHKDPKAVAT 180  
QY 181 DSHRMSQRLITLIDNTSADLMVVLPSKSLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240  
DB 181 DSHRLSQKLLTEKNSDDFDVVPISRLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240  
QY 241 YTRLLEGNYPDTDRLLMTTEFTEVFNQSLRHAMERAFISNATQNGTVKLEITQNHIS 300  
DB 241 YTRLLEGNYPDTDRLLMTTEFTEVFNQSLRHAMERAFISNATQNGTVKLEITQNHIS 300  
QY 301 AHVNSPEVGVKNEDDIDVSGSGLTISFNPTLYLESKAIKSETVKHFLSPVPRFTLT 360  
DB 301 AHVNSPEVGVKNEDDIDVSGSGLTISFNPTLYLESKAIKSETVKHFLSPVPRFTLT 360  
QY 361 PGDEEESFIQLITPVRTN 378  
DB 361 PADTDEDFMQLITPVRTN 378

## RESULT 5

US-09-815-242-10907  
Sequence 10907, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert

```

; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10907
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10907

```

[illegible]

RESULT 6  
US-10-282-122A-56984  
; Sequence 56984, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu

```

; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haeelbeck, Robert
; APPLICANT: Ohsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56984
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-10-282-122A-56984

```

Query Match	46.4%	Score 864.5	DB 15	Length 376
Best Local Similarity	45.9%	Pred. No.'5.7e-63		
Matches	173	Conservative 77	Mismatches 126	Indels 1
QY	2	IQFSINRTLFIHALNTTKRAISTKNAIPILUSSIKIEVTSSTGVILTNGNGOISIENTIPVS	61	
Db	1	MKUTVKRSVFOQLQTVQRAISSKTTIPILTVKIVLSGEDGLSLTGNADISIESFLSKD	60	
QY	62	NENAGLLITSPGAILLEAGFFINIISSLPDISINVKEIEOHQVVLTSKGSEITLTKGKD	121	
Db	61	DEKAQMTIERTSGIVLQSRFFGIIIRKLPEDMETMEVLNNOVAITSKGADFVNGLD	120	
QY	122	QYPRLOEVSTENPLIIKTKLLKSIIAETAPAAQLQBSRPILTVGHVLSNHKDFKAVATD	181	
Db	121	NYPHLEVIDITQONMKLPVHLTKIISSETGFAVSMHESRPILTVGHVLFLENQK-LLA	179	
QY	182	SHRMSQRLITLTDNTSADLMVLPKSLIRFSAVFTDDIETVEVFPSPQILTFRSEHSFY	241	
Db	180	SHRLSORVPIPTQAEVDFNIVIPGKSLTELRSGLTNEEEMVEISIMENQVLFKTEITMYF	239	
QY	242	TRLLEGNYPTDRLMLTFEFYEVFNFTQSLRHAMERAFILSNATQGTVKLEITQNHISA	301	
Db	240	SRLLLEGNYPTNRLIPTSHNTQIEFVVPPELLSAIERASLLSHBGRNNIVELSISSPDSV	299	
QY	302	HYNSPEVGVKNEDDIVSOGSDLTISFNPTYLIESLKAIKSETVKIHLFSPVRPFTLT	361	
Db	300	YGNISPEIGKVEBALNYENVSGEALDISFNPDYMKDALRAFGDMNITVKFLSRPFTLE	359	
QY	362	GDEEFSFIQLITPVRTN	378	



```

Db      61 QVDGEEIVETETGVVLPGRFFVDIIKKLPGEVKLSTNEQOTLTITSGHSEFNLSGLD 120
QY      120 VDOYPRLOEVSTENPLILTKLKSIIAETAFASLOESRPILTGVHIVLSNHKDFKAVA 179
Db      121 PDQYPLPEVSRDAILQLSVKVLKNIIAQTNFAVSTSETRPVLTVGNWLIQDN-ELICTA 179
QY      180 TDSHRMSQRLITLTDNTSADLMVLPKSLRREFSAVFTDDIETVEVFFSPQILFRSEHIS 239
Db      180 TDSHRLAVRKLQLEDESENKNVLPKALSELNKMDSDEDDIDIFFASQVLFVGVNIN 239
QY      240 FYTRLEGNYPDTRLLMTETFEVNTQSLRHAMERAFILSNATQNGTVKLEITQNH 299
Db      240 FISRLLEGHPDTRLPFENYEIKLGINNGDFYHAIDRASLLAREGGNNVILKSTGNELV 299
QY      300 SAHVNSPEVGKVEDLDIVSQSSDLTISFNPTYLIESLKAIKSETVKIHLSPVRPFTL 359
Db      300 ELSTSTPEIGTVKEEVNANDVEGNNKISFNSKYMMDALKADINDVEVEFFGTWKPFTL 359
QY      360 TPGDEEESFIQLITPVRT 377
Db      360 KPKD-DDSVTQLILPIRT 376

```

RESULT 9

```

US-10-282-122A-71908
; Sequence 71908, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71908
; LENGTH: 377
; TYPE: PRT
; ORGANISM: Staphylococcus haemolyticus
US-10-282-122A-71908

```

```

Query Match      39.5%; Score 735.5; DB 15; Length 377;
Best Local Similarity 39.6%; Pred. No. 2.7e-52;
Matches 150; Conservative 90; Mismatches 134; Indels 5; Gaps 4;

QY      1 MIQFSINRFLTHALNTTKRAISTKNAIPILSSIKIEVTSTGTTLTSGNQISIENTIP- 59
Db      1 MMEFTIRRDYFINQLNDTLKALSPPTTPILTGIKIDAKDNEVILTGSDEISISITIPK 60
QY      60 -VSNENAGLLITSPGAILLEASFFINISSLPPDISINVKIEBHQHVLTSGKSEITLKGK 118
Db      61 QVDGED1-VTISSETGSVVLPGRFFVDIIKKLPKGVKLSNEQFOTLTITSGHSEFNLSGL 119
QY      119 DVQYPRLOEVSTENPLILTKLKSIIAETAFASLOESRPILTGVHIVLSNHKDFKAV 178
Db      120 DPQYPLPEVSRDAILQLSVKVLKNIIAQTNFAVSTSETRPVLTVGNWLIQDN-ELICT 178
QY      179 ATDSHRMSQRLITLTDNTSADLMVLPKSLRREFSAVFTDDIETVEVFFSPQILFRSEHI 238
Db      179 ATDSHRLAVRKLQLEDESENKNVLPKALSELNKMDSDEDDIDIFFASQVLFKGVN 238
QY      239 SPYTRLEGNYPDTRLLMTETFEVNTQSLRHAMERAFILSNATQNGTVKLEITQNH 298
Db      239 FISRLLEGHPDTRLPFENYEIKLGINNGDFYHAIDRASLLAREGGNNVILKSTGNDV 298
QY      299 ISAHVNSPEVGKVEDLDIVSQSSDLTISFNPTYLIESLKAIKSETVKIHLSPVRPFT 358
Db      299 VELSTSTPEIGTVKEEVNANDVEGNNKISFNSKYMMDALKADINDVEVEFFGTWKPFI 358
QY      359 LTPGDEEESFIQLITPVRT 377
Db      359 LKP-KEDDSVTQLILPIRT 376

```

RESULT 10

```

US-09-815-242-12700
; Sequence 12700, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12700
; LENGTH: 377
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12700

```

Query Match	39.2%;	Score 730.5;	DB 9;	Length 377;
Best Local Similarity	39.3%;	Pred. NO. 7e-52;		
Matches '149; Conservative	90;	Mismatches 135;	Indels 5;	Gaps 4;
QY	1	MIQFSNRFLTHALNTTKRAISTKNAIPILASIKLEVSTGTVLTGNSGQISIENTIPV	60	
DB	1	MMEFTIKRDYFTQLMDTLKALSPRTLTPILGIKIDAKEHEVILTGSEISIEITPK	60	
QY	61	SNENAGLL-ITSPGAILLEASPFINIISLLPDISINVKBEIQHQVVLTSKGSSEITLKGKD	119	
DB	61	TVDGEDIVNISETGSVVLPGRFVVDIIKKLPGKDVKLSTNEQOTLITSGHSEFNLSGLD	120	
QY	120	VDOYPELQSVSTENPALTKLLKSIIAETAAASLOESRPILTGVHIVLSNHKDPKAVA	179	
DB	121	PQOYPLLPVQSRDDAQLSVKVLKNVIQATNFVSTSETRPVLVTGNWLIQEN-ELICTA	179	
QY	180	TDSHRMSQRLITIDNTSADMLVVLPSKLSREFSAVFTDDIETVEVFPSPQLFRSEHIS	239	
DB	180	TDSHRLAVRKLQLEDVSEKNKVIIPGKALAEINKIMSDNEEDIDIFFASQVLFKGVNVN	239	
QY	240	FYTRLLGEGYPTDRLLMTETETEVVFNVTQSLRHAMERAFLSNATQNGTVKLEITQNHI	299	
DB	240	FISRLLEGYPTDRFLFPENYIKLSINDGEFYHAIDRASLLAREGGNNVIKLSTGDDVV	299	
QY	300	SAHVNSPEVGKYNEDLDIVSQSGSDLTITFNPNTYLIESLAKTSEVKIHFILSPVPFTL	359	
DB	300	ELSSITSPELGTKEEVANDVEGSLKISFNSKYMMDALKADNDEVEVEFFGTMKPFIL	359	
QY	360	TP-GDEESFIQLITPVRT	377	
DB	360	KPKGD-DSVTOLILPIRT	376	

```

RESULT 12
US-10-282-287-10
; Sequence 10, Application US/10282287
; Publication No. US20030129633A1
; GENERAL INFORMATION:
; APPLICANT: O'Donnell, Michael E.
; APPLICANT: Zhang, Dan
; APPLICANT: Whipple, Richard
; TITLE OF INVENTION: DNA REPLICATION PROTEINS OF GRAM POSITIVE BACTERIA AND
; TITLE OF INVENTION: THEIR USE TO SCREEN FOR CHEMICAL INHIBITORS
; FILE REFERENCE: 22221/1002
; CURRENT APPLICATION NUMBER: US/10/282,287
; CURRENT FILING DATE: 2002-10-28
; PRIOR APPLICATION NUMBER: US/09/235,245
; PRIOR FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/074,522
; PRIOR FILING DATE: 1998-01-27
; PRIOR APPLICATION NUMBER: 60/093,727
; PRIOR FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 377
; TYPE: PR1
; ORGANISM: Staphylococcus aureus
US-10-282-287-10

Query Match          39.2%; Score 730.5; DB 14; Length 377;
Best Local Similarity 39.3%; Pred: No. 7e-52;
Matches 149; Conservative 90; Mismatch 135; Indels 5; Gaps 4;

Qy      1  MIOFSINRFLTHALNTKRAISTKNAIFILSIKIEVTSTGVTLTSGNGQISIENTIPV 60
      : : : : : | | | | | : : : : : | | | : : | | | : : | | |
Db      1  MMEFTIKRDYFTQLNDTLKALSPRTLILTCIKIDAKEHEVILTGSDSEISIEITIPK 60

Qy      61  SHENAGLL-ITSPGAILLEASFPFINIISSLPDISINVKIEHQHVLTSGKGEITUKGD 119
      : : : : : | | | : : | : : : : : | : : : : : | : : : : :
Db      61  TVDGEDIVNISETGVSVLPGRFVDIIKKLPKDVKLSTNEQFQTLITSGHSEFNLSGLD 120

```





Db 121 PQYPLLQVSRDDAIQUSVKLVKNVIAQTNAVSTSETRPVLTVGNWLIQEN-ELICTA 179  
QY 180 TDSHMSORLIATDNTSADLMVLPKSLRSEFSAVFTDDIETVEVFFSPSQILFRSEHIS 239  
Db 180 TDSHRLAVRKLEQEDVSENKNVIFCKALAEINKIMSDNEEDIDIFFASQVLPKVGNNV 239  
QY 240 FYTRLLEGNYPTDRLMTPEFTEVFNQSRHMERAFNISNATQNGTVKLEITQNH 299  
Db 240 FISRLLEGHYPTDRLFPENYEIKLSIDNGEYHAIDRASLLAREGGNNVILKSTGDDVV 299  
QY 300 SAHVNSPEVGVKNEDLDIVSQSSDLTISFNPTYLIESKATKSETVKIHFELSPVRPTL 359  
Db 300 ELSSTSPETGTVEEVANDVVEGSLKISFNKYMMDALKAIDNDEVEVEFGTMRKPFIL 359  
QY 360 TP-GDEESFIQILTPVRT 377  
Db 360 PKGD--DSVTQLILPIRT 376

## RESULT 15

US-09-815-242-5669  
; Sequence 5669, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011a  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5669  
; LENGTH: 376  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-5669

Query Match 38.9%; Score 725.5; DB 9; Length 376;  
Best Local Similarity 39.2%; Pred. No. 1.8e-51;  
Matches 148; Conservative 90; Mismatches 135; Indels 5; Gaps 4;

QY 2 IQFSNRTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVLTVGSGNGQISIENTIPVS 61  
Db 1 MEFTTKRDYFTQLNDTLKALSPRTLPIITGKIDAKEHEVILTVGSGSEISIEITIPKT 60  
QY 62 NENAGL-ITSPGAILLEASFFINISSLIPDISINVKIEQHQVVLTVGSKSEITLKGXDV 120  
Db 61 VGDGDIVNISGTVLPGRFFVDIIKKLPKGVKLVSTNEQFQTLTSGHSEFNLSGLDP 120  
QY 121 DOYPRLOEVSTENPILKTLKLSIAETAFAASIQESRPILTGVHVLVSNHOKFAVAT 180

Db 121 DOYPLLQVSRDDAIQUSVKLVKNVIAQTNAVSTSETRPVLTVGNWLIQEN-ELICTAT 179  
QY 181 DSHMSORLIATDNTSADLMVLPKSLRSEFSAVFTDDIETVEVFFSPSQILFRSEHISF 240  
Db 180 DSHRLAVRKLEQEDVSENKNVIFCKALAEINKIMSDNEEDIDIFFASQVLPKVGNNV 239  
QY 241 FYTRLLEGNYPTDRLMTPEFTEVFNQSRHMERAFNISNATQNGTVKLEITQNHIS 300  
Db 240 ISRLLEGHYPTDRLFPENYEIKLSIDNGEYHAIDRASLLAREGGNNVILKSTGDDVV 299  
QY 301 AHVNSPEVGVKNEDLDIVSQSSDLTISFNPTYLIESKATKSETVKIHFELSPVRPTLT 360  
Db 300 LSSTSPETGTVEEVANDVVEGSLKISFNKYMMDALKAIDNDEVEVEFGTMRKPFILK 359  
QY 361 P-GDEESFIQILTPVRT 377  
Db 360 PKGD--DSVTQLILPIRT 375

## RESULT 16

US-10-282-122A-60333  
; Sequence 60333, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 60333  
; LENGTH: 381  
; TYPE: PRT  
; ORGANISM: Listeria monocytogenes  
US-10-282-122A-60333

Query Match 37.7%; Score 703; DB 15; Length 381;  
Best Local Similarity 38.8%; Pred. No. 1.3e-49;  
Matches 148; Conservative 93; Mismatches 134; Indels 6; Gaps 5;



```
Query Match 31.5%; Score 586; DB 15; Length 334;
Best Local Similarity 37.9%; Pred. No. 5.3e-40;
Matches 127; Conservative 74; Mismatches 126; Indels 8; Gaps 7;

QY 49 NQGIS-IENTIPVSNENAGLL-ITSPGAILLEASFFINIISLPDISINVKIEHQHVVL 106
Db 1 NSDSIIIESFIPLEKEGKLLVDVKRPGSIVLQARFSEIVKLPQQTVEIETEDNFLI 60

QY 107 TSGKSEITLKGVDQVYPRLOEVSTENPILKTKLLKSIATFAASLOESRPILTVGH 166
Db 61 RSGHSEFRNLGNLADEYPRLPQIEENVFQIPADLLKTVIRQTVFAVSTSETRPILTVGN 120

QY 167 IVLSNHKDFKAVATSDHRMSQRLITLDNTSADLM--VVLPSKSLREFSAVFTDD-IETVE 223
Db 121 WKV-EHSELVCTATDSHRLAMRKVKIIIESENEVSNVIVPGKSLNELSKIIIDGHPVD 179

QY 224 VFFSPSQILFRSEHISFYTRLLEGNYPTDRLMLTTEFETEVVFNQSLRHAMERAFILSN 283
Db 180 IVMTANQVLFKAEHLFFSRLLDGNYPETARLIPTESKTTMIVNAKEFLQAIIDRASLLAR 239

QY 284 ATQNGTVKL-EITQNHISAHVNSPEVGKVNEDLDIVSOGSDLTISFNPTYLIESLKAIK 342
Db 240 EGRNVVVKLTLPGLMELISSISPEIGKVTQLOQTESLEGEELNISFSKAYMMDALRALD 299

QY 343 SETVKIHLFSVRPPTLTPGDEEESFIQIITPVRT 377
Db 300 GTDIIQISFTGMRPFLLRP-LHTDSMLQLILPVRT 333

RESULT 19
US-10-671-419-174
; Sequence 174, Application US/10671419
; Publication No. US20040038290A1
; GENERAL INFORMATION:
; APPLICANT: O'Donnell, Michael E.
; APPLICANT: Yuzhakov, Alexander
; APPLICANT: Yurieva, Olga
; APPLICANT: Jeruzalmi, David
; APPLICANT: Bruck, Irina
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT
; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND
; FILE REFERENCE: 22221/1030
; CURRENT APPLICATION NUMBER: US/10/671,419
; PRIOR FILING DATE: 2003-09-25
; PRIOR APPLICATION NUMBER: US/09/716,964A
; PRIOR FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 60/143,202
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 08/823,407
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 09/057,416
; PRIOR FILING DATE: 1998-04-08
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 174
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-10-671-419-174

Query Match 31.5%; Score 586; DB 15; Length 334;
Best Local Similarity 37.9%; Pred. No. 5.3e-40;
Matches 127; Conservative 74; Mismatches 126; Indels 8; Gaps 7;

QY 49 NQGIS-IENTIPVSNENAGLL-ITSPGAILLEASFFINIISLPDISINVKIEHQHVVL 106
Db 1 NSDSIIIESFIPLEKEGKLLVDVKRPGSIVLQARFSEIVKLPQQTVEIETEDNFLI 60

QY 107 TSGKSEITLKGVDQVYPRLOEVSTENPILKTKLLKSIATFAASLOESRPILTVGH 166
Db 61 RSGHSEFRNLGNLADEYPRLPQIEENVFQIPADLLKTVIRQTVFAVSTSETRPILTVGN 120

QY 167 IVLSNHKDFKAVATSDHRMSQRLITLDNTSADLM--VVLPSKSLREFSAVFTDD-IETVE 223
Db 121 WKV-EHSELVCTATDSHRLAMRKVKIIIESENEVSNVIVPGKSLNELSKIIIDGHPVD 179

QY 224 VFFSPSQILFRSEHISFYTRLLEGNYPTDRLMLTTEFETEVVFNQSLRHAMERAFILSN 283
Db 180 IVMTANQVLFKAEHLFFSRLLDGNYPETARLIPTESKTTMIVNAKEFLQAIIDRASLLAR 239

QY 284 ATQNGTVKL-EITQNHISAHVNSPEVGKVNEDLDIVSOGSDLTISFNPTYLIESLKAIK 342
Db 240 EGRNVVVKLTLPGLMELISSISPEIGKVTQLOQTESLEGEELNISFSKAYMMDALRALD 299

QY 343 SETVKIHLFSVRPPTLTPGDEEESFIQIITPVRT 377
Db 300 GTDIIQISFTGMRPFLLRP-LHTDSMLQLILPVRT 333

RESULT 20
US-10-670-844-174
; Sequence 174, Application US/10670844
; Publication No. US20040043414A1
; GENERAL INFORMATION:
; APPLICANT: O'Donnell, Michael E.
; APPLICANT: Yuzhakov, Alexander
; APPLICANT: Yurieva, Olga
; APPLICANT: Jeruzalmi, David
; APPLICANT: Bruck, Irina
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT
; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND
; FILE REFERENCE: 22221/1030
; CURRENT APPLICATION NUMBER: US/10/670,844
; PRIOR FILING DATE: 2003-09-25
; PRIOR APPLICATION NUMBER: US/09/716,964A
; PRIOR FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 60/143,202
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 08/823,407
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 09/057,416
; PRIOR FILING DATE: 1998-04-08
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 174
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-10-670-844-174

Query Match 31.5%; Score 586; DB 15; Length 334;
Best Local Similarity 37.9%; Pred. No. 5.3e-40;
Matches 127; Conservative 74; Mismatches 126; Indels 8; Gaps 7;

QY 49 NQGIS-IENTIPVSNENAGLL-ITSPGAILLEASFFINIISLPDISINVKIEHQHVVL 106
Db 1 NSDSIIIESFIPLEKEGKLLVDVKRPGSIVLQARFSEIVKLPQQTVEIETEDNFLI 60

QY 107 TSGKSEITLKGVDQVYPRLOEVSTENPILKTKLLKSIATFAASLOESRPILTVGH 166
Db 61 RSGHSEFRNLGNLADEYPRLPQIEENVFQIPADLLKTVIRQTVFAVSTSETRPILTVGN 120

QY 167 IVLSNHKDFKAVATSDHRMSQRLITLDNTSADLM--VVLPSKSLREFSAVFTDD-IETVE 223
Db 121 WKV-EHSELVCTATDSHRLAMRKVKIIIESENEVSNVIVPGKSLNELSKIIIDGHPVD 179

QY 224 VFFSPSQILFRSEHISFYTRLLEGNYPTDRLMLTTEFETEVVFNQSLRHAMERAFILSN 283
Db 180 IVMTANQVLFKAEHLFFSRLLDGNYPETARLIPTESKTTMIVNAKEFLQAIIDRASLLAR 239

QY 284 ATQNGTVKL-EITQNHISAHVNSPEVGKVNEDLDIVSOGSDLTISFNPTYLIESLKAIK 342
Db 240 EGRNVVVKLTLPGLMELISSISPEIGKVTQLOQTESLEGEELNISFSKAYMMDALRALD 299

QY 343 SETVKIHLFSVRPPTLTPGDEEESFIQIITPVRT 377
Db 300 GTDIIQISFTGMRPFLLRP-LHTDSMLQLILPVRT 333
```

Db 240 EGRNNVVKLTLPGLMELISSISPEIGKVTBQLQTESLEGEELNISFSKATYMDALRALD 299  
QY 343 SETVKIHLSPVRPFTLTPGDEEESFIQLITPVRT 377  
Db 300 GTDIOISFTGMRPFLLRP-LHTDSMLQLILPVRT 333

## RESULT 21

US-10-671-134-174  
; Sequence 174, Application US/10671134  
; Publication No. US20040043415A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Donnell, Michael E.  
; APPLICANT: Yuzhakov, Alexander  
; APPLICANT: Yuzieva, Olga  
; APPLICANT: Jeruzalmi, David  
; APPLICANT: Bruck, Irina  
; APPLICANT: Kuriyan, John  
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT  
; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND  
; FILE REFERENCE: 22221/1030  
; CURRENT APPLICATION NUMBER: US/10/671,134  
; CURRENT FILING DATE: 2003-09-25  
; PRIOR APPLICATION NUMBER: US/09/716,964A  
; PRIOR FILING DATE: 2000-11-21  
; PRIOR FILING DATE: 1997-04-08  
; PRIOR FILING DATE: 1997-04-08  
; PRIOR FILING DATE: 1997-04-08  
; PRIOR FILING DATE: 1998-04-08  
; NUMBER OF SEQ ID NOS: 212  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 174  
; LENGTH: 334  
; TYPE: PRT  
; ORGANISM: Bacillus stearothermophilus  
US-10-671-134-174

Query Match 31.5%; Score 586; DB 15; Length 334;  
Best Local Similarity 37.9%; Pred. No. 5.3e-40;  
Matches 127; Conservative 74; Mismatches 126; Indels 8; Gaps 7;  
QY 49 NGQIS-IENTIPVSNENAGLL-ITSPGAILLEASFFINIISLPDISINVKIEBQHVVL 106  
Db 1 NSDISIESFIPLEKEGKLLVDVVRPGSIVLQARFFSEIVKLPQQTVEIETEDNFLTII 60  
QY 107 TSGKSEITLKGKVDQYPRLOEVSTENPLILKTKLSIIAETAFAASLOESRPILTVGH 166  
Db 61 RSGHSEFRNLGNLADEYPRLPQIEENVFOIPADLLKTVIRQTVFAVSTETRPILTVGN 120  
QY 167 IVLSNHKPKAVATDSHRMSORLITLNTSADLM--VVLPSKSLREFSAVFTDD-IETVE 223  
Db 121 WKV-EGHELVCVATDSHRLAMRKVIIESENEVSNVVIIPGKSLNELSKILLDGNHPVD 179  
QY 224 VFSPSQILFRSEHISFYTRLEGNYPDTDRLLMTETFEVNTQSLRHAMERAFILSN 283  
Db 180 IVMANQVLPKABHLLFFSRLLDGNYPETARLIPTESKTTMIVNAKEFLQADRASILAR 239  
QY 284 ATQNGTVKL-EITQNHISAHVNSPEVGVKNEDLDIVSQSGDLTISFNPTLYLESKAIK 342  
Db 240 EGRNNVVKLTLPGLMELISSISPEIGKVTBQLQTESLEGEELNISFSKATYMDALRALD 299  
QY 343 SETVKIHLSPVRPFTLTPGDEEESFIQLITPVRT 377  
Db 300 GTDIOISFTGMRPFLLRP-LHTDSMLQLILPVRT 333

## RESULT 22

US-10-673-098-174  
; Sequence 174, Application US/10673098  
; Publication No. US20040048309A1

; GENERAL INFORMATION:  
; APPLICANT: O'Donnell, Michael E.  
; APPLICANT: Yuzhakov, Alexander  
; APPLICANT: Yuzieva, Olga  
; APPLICANT: Jeruzalmi, David  
; APPLICANT: Bruck, Irina  
; APPLICANT: Kuriyan, John  
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT  
; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND  
; FILE REFERENCE: 22221/1030  
; CURRENT APPLICATION NUMBER: US/10/673,098  
; CURRENT FILING DATE: 2003-09-26  
; PRIOR APPLICATION NUMBER: US/09/716,964A  
; PRIOR FILING DATE: 2000-11-21  
; PRIOR APPLICATION NUMBER: 60/143,202  
; PRIOR FILING DATE: 1997-04-08  
; PRIOR APPLICATION NUMBER: 08/823,407  
; PRIOR FILING DATE: 1997-04-08  
; PRIOR APPLICATION NUMBER: 09/057,416  
; PRIOR FILING DATE: 1998-04-08  
; NUMBER OF SEQ ID NOS: 212  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 174  
; LENGTH: 334  
; TYPE: PRT  
; ORGANISM: Bacillus stearothermophilus  
US-10-673-098-174

Query Match 31.5%; Score 586; DB 15; Length 334;  
Best Local Similarity 37.9%; Pred. No. 5.3e-40;  
Matches 127; Conservative 74; Mismatches 126; Indels 8; Gaps 7;  
QY 49 NGQIS-IENTIPVSNENAGLL-ITSPGAILLEASFFINIISLPDISINVKIEBQHVVL 106  
Db 1 NSDISIESFIPLEKEGKLLVDVVRPGSIVLQARFFSEIVKLPQQTVEIETEDNFLTII 60  
QY 107 TSGKSEITLKGKVDQYPRLOEVSTENPLILKTKLSIIAETAFAASLOESRPILTVGH 166  
Db 61 RSGHSEFRNLGNLADEYPRLPQIEENVFOIPADLLKTVIRQTVFAVSTETRPILTVGN 120  
QY 167 IVLSNHKPKAVATDSHRMSORLITLNTSADLM--VVLPSKSLREFSAVFTDD-IETVE 223  
Db 121 WKV-EGHELVCVATDSHRLAMRKVIIESENEVSNVVIIPGKSLNELSKILLDGNHPVD 179  
QY 224 VFSPSQILFRSEHISFYTRLEGNYPDTDRLLMTETFEVNTQSLRHAMERAFILSN 283  
Db 180 IVMANQVLPKABHLLFFSRLLDGNYPETARLIPTESKTTMIVNAKEFLQADRASILAR 239  
QY 284 ATQNGTVKL-EITQNHISAHVNSPEVGVKNEDLDIVSQSGDLTISFNPTLYLESKAIK 342  
Db 240 EGRNNVVKLTLPGLMELISSISPEIGKVTBQLQTESLEGEELNISFSKATYMDALRALD 299  
QY 343 SETVKIHLSPVRPFTLTPGDEEESFIQLITPVRT 377  
Db 300 GTDIOISFTGMRPFLLRP-LHTDSMLQLILPVRT 333

## RESULT 23

US-10-672-638-174  
; Sequence 174, Application US/10672638  
; Publication No. US20040077012A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Donnell, Michael E.  
; APPLICANT: Yuzhakov, Alexander  
; APPLICANT: Yuzieva, Olga  
; APPLICANT: Jeruzalmi, David  
; APPLICANT: Bruck, Irina  
; APPLICANT: Kuriyan, John  
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT  
; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND  
; FILE REFERENCE: 22221/1030

```
; CURRENT APPLICATION NUMBER: US/10/672,638
; CURRENT FILING DATE: 2003-10-09
; PRIOR APPLICATION NUMBER: US/09/716,964A
; PRIOR FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 60/143,202
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 08/823,407
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 09/057,416
; PRIOR FILING DATE: 1998-04-08
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 174
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-10-673-127-174

Query Match      31.5%; Score 586; DB 15; Length 334;
Best Local Similarity 37.9%; Pred. No. 5.3e-40;
Matches 127; Conservative 74; Mismatches 126; Indels 8; Gaps 7;

QY 49 NQGIS-IENTIPVSNENAGLL-ITSPGAILLEASFFINISSLPLDISINVKIEHQHVVL 106
DB 1 NSDISIIIESFPLEKEGKLLVDVVRPGSVIVLOARFSEIVKLPQOTVEIETEDNFLI 60
QY 107 TSGKSEITLKGKVDQYPRLOEVSTENPLILKTLKLSIIAETAFAASLOESRPILTVGH 166
DB 61 RSGHSEFRLGNLADEYPRLPQIEEENVFQIPADLLKTVIRQTVFAVSTSETRPILTVGN 120
QY 167 IVLSNHKDFKAVATDSHRMSORLITLNTSADLM--VVLPSKSLREFSAVFTDD-IETVE 223
DB 121 WKV-EHGELVCTATDSHRLAMRKVKIIESENEVSNVVIIPGKSLNELSKILLDDGNHPVD 179
QY 224 VFFSPSQILFRSEHISFYTRLLEGNYPDTDRLLMTTEFETEVVNTQSLRHAMERAFILSN 283
DB 180 IVMANQVLFKAEHLFFSRLLDGNYPETARLIPTESKTIWVNAKEFLQALDRASLLAR 239
QY 284 ATQNGTVKL-EITQNHISAHVNSPEVGKVNEDLDIVSGSDLTISFNPTYLIESLKAIK 342
DB 240 EGRNVVVKLTLPGGMLEISSISPEIGKVTEQLQTESLEGEELNIFSASKYMMDALRALD 299
QY 343 SETVKIHLSPVRPTLTPGDEESFQILITPVRT 377
DB 300 GTDIQISFTGAMRPFLLRP-LHTDSMLQLILPVRT 333

RESULT 24
US-10-673-127-174
; Sequence 174, Application US/10673127
; Publication No. US20040081995A1
; GENERAL INFORMATION:
; APPLICANT: O'Donnell, Michael E.
; APPLICANT: Yuzhakov, Alexander
; APPLICANT: Yurieva, Olga
; APPLICANT: Jeruzalmi, David
; APPLICANT: Bruck, Irina
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT
; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND
; FILE REFERENCE: 22221/1030
; CURRENT APPLICATION NUMBER: US/10/673,127
; PRIOR FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: US/09/716,964A
; PRIOR FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 60/143,202
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 08/823,407
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 09/057,416
; PRIOR FILING DATE: 1998-04-08
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 174
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Bacillus stearothermophilus
US-10-673-127-174

Query Match      31.5%; Score 586; DB 15; Length 334;
Best Local Similarity 37.9%; Pred. No. 5.3e-40;
Matches 127; Conservative 74; Mismatches 126; Indels 8; Gaps 7;

QY 49 NQGIS-IENTIPVSNENAGLL-ITSPGAILLEASFFINISSLPLDISINVKIEHQHVVL 106
DB 1 NSDISIIIESFPLEKEGKLLVDVVRPGSVIVLOARFSEIVKLPQOTVEIETEDNFLI 60
QY 107 TSGKSEITLKGKVDQYPRLOEVSTENPLILKTLKLSIIAETAFAASLOESRPILTVGH 166
DB 61 RSGHSEFRLGNLADEYPRLPQIEEENVFQIPADLLKTVIRQTVFAVSTSETRPILTVGN 120
QY 167 IVLSNHKDFKAVATDSHRMSORLITLNTSADLM--VVLPSKSLREFSAVFTDD-IETVE 223
DB 121 WKV-EHGELVCTATDSHRLAMRKVKIIESENEVSNVVIIPGKSLNELSKILLDDGNHPVD 179
QY 224 VFFSPSQILFRSEHISFYTRLLEGNYPDTDRLLMTTEFETEVVNTQSLRHAMERAFILSN 283
DB 180 IVMANQVLFKAEHLFFSRLLDGNYPETARLIPTESKTIWVNAKEFLQALDRASLLAR 239
QY 284 ATQNGTVKL-EITQNHISAHVNSPEVGKVNEDLDIVSGSDLTISFNPTYLIESLKAIK 342
DB 240 EGRNVVVKLTLPGGMLEISSISPEIGKVTEQLQTESLEGEELNIFSASKYMMDALRALD 299
QY 343 SETVKIHLSPVRPTLTPGDEESFQILITPVRT 377
DB 300 GTDIQISFTGAMRPFLLRP-LHTDSMLQLILPVRT 333

RESULT 24
US-10-673-127-174
; Sequence 174, Application US/10673127
; Publication No. US20040081995A1
; GENERAL INFORMATION:
; APPLICANT: O'Donnell, Michael E.
; APPLICANT: Yuzhakov, Alexander
; APPLICANT: Yurieva, Olga
; APPLICANT: Jeruzalmi, David
; APPLICANT: Bruck, Irina
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT
; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND
; FILE REFERENCE: 22221/1030
; CURRENT APPLICATION NUMBER: US/10/673,127
; PRIOR FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: US/09/716,964A
; PRIOR FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 60/143,202
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 08/823,407
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 09/057,416
; PRIOR FILING DATE: 1998-04-08
; NUMBER OF SEQ ID NOS: 212
```

[illegible]

RESULT 26

```

US-10-673-119-174
; Sequence 174, Application US/10673119
; Publication No. US20040110210A1
; GENERAL INFORMATION:
; APPLICANT: O'Donnell, Michael E.
; APPLICANT: Yuzhakov, Alexander
; APPLICANT: Yuriyeva, Olga
; APPLICANT: Jeruzalmi, David
; APPLICANT: Bruck, Irina
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT
; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND
; TITLE OF INVENTION: USE THEREOF
; FILE REFERENCE: 22221/1030
; CURRENT APPLICATION NUMBER: US/10/673,119
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: US/09/716,964A
; PRIOR FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 60/143,202
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 08/823,407
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 09/057,416
; PRIOR FILING DATE: 1998-04-08
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 174
; LENGTH: 334
; TYPE: FRT
; ORGANISM: Bacillus stearothermophilus
US-10-673-119-174

```

```

QY 224 VFFSPQILPRSEHISFYRLLEGNYPDDBLLMTFETEVVNTQSLRHAMERAFILSN 283
Db 180 IMTANQVLFKAELHFFSRLLDGNYPETARLIPTESKTTMIVNAKEFLQAI DRASLLAR 239
QY 284 ATONGTKVL-RITQNHISAHVNSPEVGKVNEDLDIVSOGSDLTISNPTVLIESLKAIK 342
Db 240 EGRNNVKKTLTLPQGMLEISISFEIGKVTQLQTESLEGELNLSFSAKYMMDALRAUD 299
QY 343 SETVKKHFLSPVRPFTLTTPGDEESFIQLITPVRT 377
Db 300 GTDIQISFTGAMRPFLLRP-LHTDSMLQLILPVRT 333

RESULT 27
US-10-671-207-174
; Sequence 174, Application US/10671207
; Publication No. US20040197796A1
; GENERAL INFORMATION:
; APPLICANT: O'Donnell, Michael E.
; APPLICANT: Yuzhakov, Alexander
; APPLICANT: Yurieva, Olga
; APPLICANT: Jeruzalimi, David
; APPLICANT: Bruck, Irina
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT
; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND
; TITLE OF INVENTION: USE THEREOF
; FILE REFERENCE: 22221/1030
; CURRENT APPLICATION NUMBER: US/10/671,207
; CURRENT FILING DATE: 2003-09-25
; PRIOR APPLICATION NUMBER: US/09/716,964
; PRIOR FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 60/143,202
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 08/823,407
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 09/057,416
; PRIOR FILING DATE: 1998-04-08
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 174
; LENGTH: 334
; TYPE: PR
; ORGANISM: Bacillus stearothermophilus
US-10-671-207-174

```

Query Match	31.5%;	Score 586;	DB 17;	Length 334;
Best Local Similarity	37.9%;	Pred. No. 5.3e-40;		
Matches	127;	Conservative 74;	Mismatches 126;	Indels 8; Gaps 7;
QY	49	NGOIS-IENTYPSVNEAGLL-ITSPGALLLEASFINIISLPDISINVKETEHOQVUL	106	
Db	1	NDISIIIESFIDKEGKLLVDVKGPGSVLQARFSEIVKKLPQOTVBIETEDNFLI	60	
QY	107	TGKGEITLKGKDVOQYPRLOEVSTENPLILTKLKSIAETAFAASLQESRPILTGVH	166	
Db	61	RSCHSEFRNLGNADYPRLPOIEENVFOIPADLLKTIVQTVAVSTSETRPILTVGN	120	
QY	167	IYLSNHKDFKAVATDSHMSORLITLTDNTSADLM--VVLPSKSLREFSFAVFTDD-IETVE	223	
Db	121	WKV-BHEGLVCTATDUSHLRMKVKIIIESENVSVNVVPGKSLNELSKIIIDGNGHPVD	179	
QY	224	VFFSPQILFRSEHTSFYTRLEGNYPTDRLIMTEFEVFNFTQSLRHAMERAPLISN	283	
Db	180	IYMTANQVLFKAEHLFFSRLLDGNYPETARLIPETSKTTIMVNAKEFIQAIDRASLLAR	239	
QY	284	ATQNGTVKL-EITQNHISAHVNSPEVGKNEDLDIVSOSGSLTITSFNPTYLIESLKATK	342	
Db	240	EGRNVNVKLTWIPGGMLEISISPIEGKVTQLOTESLEGEELNISFSKAYMMMDALRAD	299	
QY	343	SETVKIHLFSPVRPTTLTPGDEESFIQLITPVRT	377	

Db 300 GTDIOISFTGMRPFLLRP-LHTDSMLQLILPVRT 333

US-10-282-122A-52020

Sequence 52020, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Lianguo

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: Patent in version 3.1

SEQ ID NO 52020

LENGTH: 366

TYPE: PRT

ORGANISM: Clostridium acetobutylicum

US-10-282-122A-52020

Query Match 24.6%; Score 457.5; DB 15; Length 366;

Best Local Similarity 29.9%; Pred. No. 2.6e-29;

Matches 112; Conservative 91; Mismatches 159; Indels 13; Gaps 7;

QY 2 IQFSNRTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGNSGQISIENTIPVS 61

Db 1 MAFICEKNILQRAITQAQAVTKSTMPVLQGLMSVQNNELTIGSDIDLSIETKINVE 60

QY 62 NENAGLLITSPGAILLEAFNIISSLPDISINVKIEHQHVLTSKGSEITLKGKQVD 121

Db 61 -----VLEEGKVLDARLLSEIRKLPNSKVEIQTENNVCVEITCNKSKLTVLYLNP 113

QY 122 QYPRLOEVSTENPLILTKLLKSIIAETAFAASLQESRILTCVHVLNHNKDFKAVATD 181

Db 114 DFPSPLEIDENSIFKINQKTLTKMTIKGTIFAQDETRPILTVGLVFEIKDSK-LNLVAID 172

QY 182 SHRMSORLITLNDTSADLMVLPKSLREFSAVFTDDIETVEVFSPSOILFRSEHISPY 241

Db 173 GYRLAURSQYIDNETS-INAVIPGKTLNEVIKILEDGD-VNITFTSNHILFNLGNTKII 230

QY 242 TRLEGNYPDTRLLMTFETFEVNTQSLRHAMERAFLLISNATONGTVKLEITQNHISA 301

Db 231 SRLLEGEFIKYNISIIPEEYNLNI VARKBELDCIERASIMAKOGNINIKLDI-EDDVMY 289

QY 302 HVNSPEVGVKVNEDLDIVSQSGDLTISFNPTVLIESKAIKSETVKIHFLSVPRFTLTP 361

Db 290 ITSNSQLGNVREINIIQ-GQPLKIAFNKYLLIDVLKIMNQEEIVMFMSSISFCIIKN 348

QY 362 GDEESFTQLITPVR 376

Db 349 KENDDS-TYLILPVR 362

RESULT 29

US-10-282-122A-53348

Sequence 53348, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Lianguo

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: Patent in version 3.1

SEQ ID NO 53348

LENGTH: 402

TYPE: PRT

ORGANISM: Clostridium difficile

US-10-282-122A-53348

Query Match 24.0%; Score 448; DB 15; Length 402;

Best Local Similarity 28.8%; Pred. No. 1.8e-28;

Matches 107; Conservative 96; Mismatches 153; Indels 16; Gaps 7;

QY 7 NRTLFHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGNSGQISIENTIPVSNENAG 66

Db 40 QKILANRIGIAQAKNGKTTIELKGLISTEEQKLTGYDAEIGIETVYQAE----- 94

QY 67 LLITSPGAILLEAFNIISSLPDISINVKIEHQHVLTSKGSEITLKGKQVDQYPL 126



Db 95 --LIEKGDVVVDARLFGDIIRKLPDSFVEIETDSENNIYVNCVNSRPFKIGYAAKEFPKL 152  
QY 127 QEVSTENPLIKTKLKJIAETAFAASLQESRPILTG--VHIVLSNKHDKFAVATDSHR 184  
Db 153 PELNEEDLYSIPQEIILKNMKIQTIVFALSQDQTPVLMGELLEIV---DRNLNLVAIDGR 209  
QY 185 MSORLTLNDSADLMVLPSPKSLREFSAVFTDDIETVEFFSPSOILFRSEHISFYTRL 244  
Db 210 LAVKSCSVDSLNIKVIIPGKTLIDVNSLLSGE-DNVKVGFEKNAFIINDTKIITRL 268  
QY 245 LEGNYPDTDRLLMTETETEVNTQSLRHAMERAFILSNATQGTGKLEITQNHISAHVN 304  
Db 269 LEGDFIDYKLLPRENSRVKUNTKELLSQSEKNNLIKLSI-RDKWMAITS 327  
QY 305 SPEVGVKNEDLIVSQSGDLTISFNPTYLIESLKAIKSETVKIHLFSLPVRPFTLTPGDE 364  
Db 328 NTEKGNVYEVEI-DLDGYLDIAFNSRYFIEGLKNIDNEEFIBETVNVNFCIKPTD- 385  
QY 365 EESFQIOLTPVR 376  
Db 386 DVNIYIYLLPVR 397

## RESULT 30

US-10-282-122A-52377  
; Sequence 52377, Application US/10282122A  
; Publication No. US20040029129A1

## GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

## TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA 034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 52377

; LENGTH: 375

; TYPE: PRT

; ORGANISM: Clostridium botulinum

US-10-282-122A-52377

Query Match 23.7%; Score 442; DB 15; Length 375;  
Best Local Similarity 29.4%; Pred No. 5.2e-28;  
Matches 113; Conservative 79; Mismatches 165; Indels 28; Gaps 8;  
QY 2 IQPSINRTLFHALMTTKRAISTKNAIPILSSIKIEVTSTGVTLTSGNSGOISIENTIPVS 61  
Db 9 MKITCQKNILLEGISIVQKAVTGKSTLPILSGILIRANKNELVLTGSDWDLSTETKVKAN 68  
QY 62 NENAGLLTSCAILLEAFINIISSIPDISINVKEIQHVLTSGKSEITLKGKVD 121  
Db 69 -----LLEGTIVLDARLFGEIIRKPNLDLIEINTLSDNSIEIICQNSRNLTYMNP 121  
QY 122 QYPRLQEVSTENPLIKTKL-----LKSIIATAFAASLQESRPILTGVIHIVLSNKH 173  
Db 122 EFP-----NPPINENMIFSGESKLNKWKGTIFATAQDETRPILTGVLFOIKD-K 172  
QY 174 DFKAVATDSHRMSQRLITLNTSADLMVLPSPKSLREFSAVFTDDIETVEFFSPSOILF 233  
Db 173 MLNLVALDGYRLALRSEVVDNDNT-INAVIPGKTLNEVSKILEEDNENNVNITFTPHILF 231  
QY 234 RSEHISFYTRLLEGNYPTDRLMLTETETEVNTQSLRHAMERAFILSNATQGTGKLE 293  
Db 232 SIGETKIISRLLEGFEFTSYKSIIPPEFNKIIAKRSELLNIESIERASLMAKSGNTNLVKPD 291  
QY 294 ITONHISAHVNSPEVGVKNEDLIVSQSGDLTISFNPTYLIESLKAIKSETVKIHLFSLP 353  
Db 292 FSDDKIVITNS-QLGWVREELKVVLO-GEDLQIAFNSKYLILDVLKTMEDNEVLEFSSS 349  
QY 354 VRPFTLTPGDEEESFQIOLTPVR 378  
Db 350 VSP-CIIKNTEINNCTVILVLEPRLN 373

## RESULT 31

US-10-282-122A-61654

; Sequence 61654, Application US/10282122A

; Publication No. US20040029129A1

## GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

## TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA 034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308



```

QY      2 IQFSINRFLFIHALNTKKAISTKKAIPILSSIKIEVTSTGVTLTGSGNQISIENTIPVS 61
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db     13 LTRFLRESFADAVSWNNKLPARPAPVPLSGVLTGSDNGLTGTGDFYVSAAQV--- 69
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
      Matches 106; Conservative 82; Mismatches 163; Indels 60; Gaps 11;
      Best Local Similarity 25.8%; Pred No. 1.7e-21;

```

```

QY 62 NENAGLLITSPGAILLEASFFINIISLSPDISINVKIEHQHVLTSGKSEITLKGKQVD 121
Db 70 ----GAEIVSPGSLVSGRLSDITRALPNKQVDV-HVEGNNRVALTCGNARFSLTMPVE 124
QY 122 QYPRLOEVSTENPLIKTKLLKSIITAEATAFAASLOESRILTVGHVILSNHDKPKAVATD 181
Db 125 DYPTLTPEETGL-LPAELFAEALISQVAIAAGRDITLPMLTGIRVILGETVVLAA-ATD 182
QY 182 SHRMSORLITLNDTADL--MVVLPKSLREFS-----AVFTDDIETV 222
Db 183 RFLAVRELKWSASSPDIEAAVLVPAKTLAEAAKAGIGSDVRLSLGTCGPGVKDGLGI 242
QY 223 EVFFSPQILFRSEHISFYTRLEGNYPDTDRLLMTEFETEVVFNQSLRHAMERAFILIS 282
Db 243 SGNKRST-----TRLLDAEFKFRQLLPTEHTAVATMDVAELIEAIKLVALV- 290
QY 283 NATONGTVKLEITQNHISAHVNSPEVGVKNEDLDIVSOGSDLTISFNPTYLIESLKAIK 342
Db 291 -ADGAQVMEFADGSRVLSAGADDVGRABEDL-VVDYAGEPLTIAFNPTYLTDLGLSLR 348
QY 343 SETVKIHFLSPVRPFTLP--GDE-----EESFIQIOLITPVR 376
Db 349 SERVSGFTTAGKPAALLRPVSGDDRPVAGLNGNGPFPVAVSTDYVYLLMPVR 399

RESULT 35
US-10-476-597-143
; Sequence 143, Application US/10476597
; Publication No. US20040235766A1
; GENERAL INFORMATION:
; APPLICANT: Bullard, James
; APPLICANT: Janjic, Nebojsa
; APPLICANT: McHenry, Charles S.
; TITLE OF INVENTION: System for Discovery of Agents that Block Yersinia pestis and
; FILE REFERENCE: RDYN.02/PCT-US
; CURRENT APPLICATION NUMBER: US/10/476,597
; PRIOR FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: US 60/290,725
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: PCT/US02/15111
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: US 60/332,644
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 143
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-476-597-143

Query Match 19.5%; Score 364; DB 17; Length 402;
Best Local Similarity 25.8%; Pred. No. 1.7e-21;
Matches 106; Conservative 82; Mismatches 163; Indels 60; Gaps 11;

QY 2 IQFSINRTLFHALNTTKRAISTKNAIPILSIKIEVTSTGVTLTGSGNGQISIENTIPVS 61
Db 13 LTFRLRESFADAVSWAKNLPARPAPVLSGLVLTGSDNGLTISGFDFEVSAAEQV--- 69
QY 62 NENAGLLITSPGAILLEASFFINIISLSPDISINVKIEHQHVLTSGKSEITLKGKQVD 121
Db 70 ----GAEIVSPGSLVSGRLSDITRALPNKQVDV-HVEGNNRVALTCGNARFSLTMPVE 124
QY 122 QYPRLOEVSTENPLIKTKLLKSIITAEATAFAASLOESRILTVGHVILSNHDKPKAVATD 181
Db 125 DYPTLTPEETGL-LPAELFAEALISQVAIAAGRDITLPMLTGIRVILGETVVLAA-ATD 182
QY 182 SHRMSORLITLNDTADL--MVVLPKSLREFS-----AVFTDDIETV 222
Db 183 RFLAVRELKWSASSPDIEAAVLVPAKTLAEAAKAGIGSDVRLSLGTCGPGVKDGLGI 242
QY 223 EVFFSPQILFRSEHISFYTRLEGNYPDTDRLLMTEFETEVVFNQSLRHAMERAFILIS 282

```

```

Db 243 SGNKRST-----TRLLDAEFKFRQLLPTEHTAVATMDVAELIEAIKLVALV- 290
QY 283 NATONGTVKLEITQNHISAHVNSPEVGVKNEDLDIVSOGSDLTISFNPTYLIESLKAIK 342
Db 291 -ADGAQVMEFADGSRVLSAGADDVGRABEDL-VVDYAGEPLTIAFNPTYLTDLGLSLR 348
QY 343 SETVKIHFLSPVRPFTLP--GDE-----EESFIQIOLITPVR 376
Db 349 SERVSGFTTAGKPAALLRPVSGDDRPVAGLNGNGPFPVAVSTDYVYLLMPVR 399

RESULT 36
US-10-282-122A-62773
; Sequence 62773, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62773
; LENGTH: 402
; TYPE: PRT
; ORGANISM: Mycobacterium bovis
US-10-282-122A-62773

Query Match 19.5%; Score 363; DB 15; Length 402;
Best Local Similarity 25.8%; Pred. No. 2e-21;
Matches 106; Conservative 81; Mismatches 164; Indels 60; Gaps 11;

QY 2 IQFSINRTLFHALNTTKRAISTKNAIPILSIKIEVTSTGVTLTGSGNGQISIENTIPVS 61
Db 13 LTFRLRESFADAVSWAKNLPARPAPVLSGLVLTGSDNGLTISGFDFEVSAAEQV--- 69
QY 62 NENAGLLITSPGAILLEASFFINIISLSPDISINVKIEHQHVLTSGKSEITLKGKQVD 121
Db 70 ----GAEIVSPGSLVSGRLSDITRALPNKQVDV-HVEGNNRVALTCGNARFSLTMPVE 124

```

```

Qy 122 QYPRLOEVSTENPLILKTKLLKSIIEAETAFASQESRPILTGWHVLSNHHKDFKAVTD 181
Db 125 DYPTLPTLPETGL-LPAELFAEASISQVATAAGRDDTLPLMTGIRVELGETVVLAA-ATD 182
Qy 182 SHRMSQRLITLNTSADL--MVLPKSLREFS-----AVFTDDIETV 222
Db 183 RFLAVRELKWSASSPDIEAALVPAKTLAAAKAGIGSDVRLSLGTGPGVKGDLGI 242
Qy 223 EVFSPQILFRSHSHISFYTRLLEGNYPDTRLLMTFETEVVNTQSLRHAMERAFILIS 282
Db 243 SGNCKRT-----TRLDAEPFKPQLLPTEHTAVATMDVAELIEAKUVALV- 290
Qy 283 NATQNGTVKLEITQNHISAHVNSPEVGNVEDLIVSQSGDLTISFNPTYLIESLKAIK 342
Db 291 -ADRGAVRMEFADGSVRLSAGADDVGRABEDL-VVDYAGEPLTIAFNPTYLTDGLSSLR 348
Qy 343 SETVKIHLSPVPFTLTP--GDE-----EESIQLITPVR 376
Db 349 SERVSFGFTAGKPALRPVSGDPRVAGLNGNGPFPFAVSTDYVLLMPVR 399

```

RESULT 37

```

US-10-282-122A-68144
; Sequence 68144, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 68144
; LENGTH: 367
; TYPE: PRT
; ORGANISM: Pseudomonas putida
US-10-282-122A-68144

```

Query Match 19.5%; Score 362.5; DB 15; Length 367;

```

Best Local Similarity 25.3%; Pred. No. 1.9e-21;
Matches 97; Conservative 96; Mismatches 164; Indels 27; Gaps 11;

Qy 2 IQFSINRRTLFHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTSGNGQISIENTIPVS 61
Db 1 MHFTIQREALLKPLQLVAGVWERRQTLFVLSNVLLVVOGQQLSLTGTDELEVGRVQLE 60
Qy 62 NENAGLLTSPCAILLEASFFINIISLP-DISINVKEIOHQVVLTSKGSEITLKGXDV 120
Db 61 EP-----AEPGEITVPARKLMDICKSLPNDALIDIK-VDEQKLVKAGRSRFTLSLTPA 113
Qy 121 DOYPRLOEVSTENPLILKTKLLKSIIEAETAFASQESRPILTGWHVLSNHHKDFKAVAT 180
Db 114 NDFPVEGSPGLTNCLEQSKRLRIETSPAMACQDDVRYVINGMLLEVSRL-TLRVST 172
Qy 181 DSHRMSQRLITLNTSADL-----MVLPKSLREFSAVFTDDIETVEVFPSPQLIFR 234
Db 173 DGHRLA-----LCSMSAPIEQEDRHQVTPRKGLELARLLTDPEGMVSIVLGQHHIRAT 227
Qy 235 SEHISFYTRLLEGNYPDTRLLMTFETEVVNTQSLRHAMERAFILISNATQNGTVKLEI 294
Db 228 TGEFTTYSKLVGDKFPDYERVLPGGDKLVVGDQALREAFSRTAILSNEKYRG-IRLQL 286
Qy 295 TQNHISAHVNSPEVGNVEDLIVSQSGDLTISFNPTYLIESLKAIKSETVKIHLSPV 354
Db 287 AAGQKIQANNPEQEEAEIEIS-VDEGSSLEIGFNVSLLDLVGMTTEQVRL-ILSDS 344
Qy 355 RPFTL--TPGDEEESFQLITPVR 376
Db 345 NSSALLOEAGNDSSYV--VMPMR 366

```

RESULT 38

```

US-10-671-403-112
; Sequence 112, Application US/10671403
; Publication No. US20040038289A1
; GENERAL INFORMATION:
; APPLICANT: O'Donnell, Michael E.
; APPLICANT: Yuzhakov, Alexander
; APPLICANT: Yurieva, Olga
; APPLICANT: Jeruzalmi, David
; APPLICANT: Bruck, Irina
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: ENZYMES DERIVED FROM THERMOPHILIC ORGANISMS THAT
; TITLE OF INVENTION: FUNCTION AS A CHROMOSOMAL REPLICASE, PREPARATION AND
; TITLE OF INVENTION: USE THEREOF
; FILE REFERENCE: 22221/1030
; CURRENT APPLICATION NUMBER: US/10/671,403
; CURRENT FILING DATE: 2003-09-25
; PRIOR APPLICATION NUMBER: US/09/716,964A
; PRIOR FILING DATE: 2000-11-21
; PRIOR APPLICATION NUMBER: 60/143,202
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 08/823,407
; PRIOR FILING DATE: 1997-04-08
; PRIOR APPLICATION NUMBER: 09/057,416
; PRIOR FILING DATE: 1998-04-08
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 112
; LENGTH: 367
; TYPE: PRT
; ORGANISM: Pseudomonas putida
US-10-671-403-112

```

Query Match 19.5%; Score 362.5; DB 15; Length 367;

```

Best Local Similarity 25.3%; Pred. No. 1.9e-21;
Matches 97; Conservative 96; Mismatches 164; Indels 27; Gaps 11;

Qy 2 IQFSINRRTLFHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTSGNGQISIENTIPVS 61
Db 1 MHFTIQREALLKPLQLVAGVWERRQTLFVLSNVLLVVOGQQLSLTGTDELEVGRVQLE 60

```

Query Match 19.5%; Score 362.5; DB 15; Length 367;



Db 228 TGEFTTSKLVGKFPDYERVLPGGDKLVVGDQALREAFRTAILSNEKYRG-IRLQL 286  
 QY 295 TONHISAHVNSPEVGKYNEDLDIVSQSDITISFNFTYLIESLKAIKSETVKIHFSPV 354  
 Db 287 AAGQKIQANNPEOEAEIEIS-VDYEGSSLEIGFNVSYLLDLVGVMTTEQVRL-IILSDS 344  
 QY 355 RPFTL--TPGDEESFIQLITPVR 376  
 Db 345 NSSALLOEAGNDSSYV--VMFMR 366

Search completed: January 28, 2005, 17:12:11  
 Job time : 150 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2005, 16:35:51 ; Search time 157 Seconds  
(without alignments)  
863.692 Million cell updates/sec

Title: US-10-048-071-28  
Perfect score: 1863  
Sequence: 1 MIQFSINTLFIHALNTTKR.....LTPGDEBSFIQLITPVRTN 378

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_23Sep04:.\*  
1: \_geneseqp1980s:.\*  
2: \_geneseqp1990s:.\*  
3: \_geneseqp2000s:.\*  
4: \_geneseqp2001s:.\*  
5: \_geneseqp2002s:.\*  
6: \_geneseqp2003as:.\*  
7: \_geneseqp2003bs:.\*  
8: \_geneseqp2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1863	100.0	378	4 AAB31944	Aab31944 Amino aci
2	1849	99.2	378	5 ABP27970	Abp27970 Streptoco
3	1849	99.2	378	5 ABG61623	Abg61623 Streptoco
4	1849	99.2	378	6 ABU46377	Abu46377 Protein e
5	1527	82.0	378	6 ABU44094	Abu44094 Protein e
6	1491	80.0	378	5 ABP27969	Abp27969 Streptoco
7	1389	74.6	378	6 ABU02772	Abu02772 S. pneumo
8	1389	74.6	378	8 ADK47890	Adk47890 Streptoco
9	1384	74.3	378	4 AAU37853	Aau37853 Streptoco
10	1384	74.3	378	6 ABU45758	Abu45758 Protein e
11	1053	56.5	380	5 ABB53301	Abb53301 Lactococc
12	864.5	46.4	376	4 AAU35314	Aau35314 Enterococ
13	864.5	46.4	376	6 ABU29060	Abu29060 Protein e
14	863.5	46.3	384	7 ADC94929	Adc94929 E. faeciu
15	862.5	46.3	376	6 ABU30083	Abu30083 Protein e
16	748.5	40.2	377	6 ABU42909	Abu42909 Protein e
17	748.5	40.2	385	5 ABP40321	Abp40321 Staphyloc
18	745	40.0	400	6 AD807184	Ad807184 Alloiococ
19	745	40.0	400	8 ADJ27077	Adj27077 Alloiococ
20	735.5	39.5	377	6 ABU43984	Abu43984 Protein e
21	730.5	39.2	377	2 AAU49071	Aau49071 Amino aci
22	730.5	39.2	377	4 AAU37133	Aau37133 Staphyloc
23	730.5	39.2	377	4 AAU37107	Aau37107 Staphyloc
24	730.5	39.2	377	4 AAB31935	Aab31935 Amino aci
25	730.5	39.2	377	5 ABB81847	Abb81847 Staphyloc

26 730.5 39.2 377 6 ABU16160  
27 730.5 39.2 377 6 ABM72805  
28 730.5 39.2 377 7 ADD26272  
29 730.5 39.2 377 7 ADD26274  
30 730.5 39.2 377 7 ABW01648  
31 725.5 38.9 376 4 RAU34173  
32 708 38.0 357 6 ADB07182  
33 703 37.7 381 5 ABB49806  
34 703 37.7 381 6 ABU32409  
35 692 37.1 379 6 ABU17704  
36 586 31.5 334 8 ADJ67944  
37 586 31.5 334 8 ADJ68156  
38 586 31.5 334 8 ADK01234  
39 586 31.5 334 8 ADJ79453  
40 586 31.5 334 8 ADJ84893  
41 586 31.5 334 8 ADM77681  
42 586 31.5 334 8 ADM66348  
43 586 31.5 334 8 ADO04401  
44 586 31.5 334 8 ADP82478  
45 541 29.0 272 6 ADA89618

## ALIGNMENTS

RESULT 1  
AAB31944  
ID AAB31944 standard; protein; 378 AA.

XX AAB31944;

XX 15-MAY-2001 (first entry)

XX Amino acid sequence of a partial dnaN (beta subunit) polypeptide.

XX dnaG; Gram positive bacteria; polC; dnaG; holA; holB; dnaX; dnaN; seb;

XX dnaG; dnaB; antibiotic; replication; cell growth; cell death;

XX bacterial infection.

XX Streptococcus pyogenes.

XX WO200109164-A2.

XX 08-FEB-2001.

XX 28-JUL-2000; 2000WO-US020666.

XX 29-JUL-1999; 99US-0146178P.

XX (UVRQ ) UNIV ROCKEFELLER.

XX O'donnell ME, Bruck I, Zhang D, Whipple R;

XX WPI; 2001-147453/15.

XX N-FSDB; AAF54744.

XX Isolated DNA molecule from a Gram positive bacterium encoding DNA replication proteins used to identify compounds which have antibiotic activity.

XX Claim 80; Page 62-63; 239pp; English.

XX The present sequence represents a partial dnaN polypeptide. The specification describes DNA molecules from Gram positive bacteria, which comprise a coding region from a polC, dnaE, holA, holB, dnaX, dnaN, seb, dnaG or a dnaB gene. These sequences encode proteins that replicate the chromosome of Gram positive bacteria. They are used for sequencing and amplification of DNA and in drug discovery to identify compounds which have antibiotic activity through interference with replication. They are used in methods for identifying compounds that are active at the level of DNA replication and result in arrest of cell growth or cell death of bacteria to treat bacterial infections in animals

XX

SQ Sequence 378 AA;

Query Match 100.0%; Score 1863; DB 4; Length 378;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-156;  
 Matches 378; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIQFSINRTLFHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNQISIENTIPV 60  
 DB 1 MIQFSINRTLFHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNQISIENTIPV 60

QY 61 SNEAGLLITSPGAILLEASFFINIISLPDISINVKIEHQVLTSGKSEITLKGKDV 120  
 DB 61 SNEAGLLITSPGAILLEASFFINIISLPDISINVKIEHQVLTSGKSEITLKGKDV 120

QY 121 DQYPRLEQVSTENPLIKTKLLSIIAETAFASLOESRPILTCVHIVLSNKHDFKAVAT 180  
 DB 121 DQYPRLEQVSTENPLIKTKLLSIIAETAFASLOESRPILTCVHIVLSNKHDFKAVAT 180

QY 181 DSHRMSQRLITLNTSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240  
 DB 181 DSHRMSQRLITLNTSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240

QY 241 YTRLLEGNYPTDRLMTTEFETEVNTQSLRHAMERAFLLISNATONGTVKLEITQNHIS 300  
 DB 241 YTRLLEGNYPTDRLMTTEFETEVNTQSLRHAMERAFLLISNATONGTVKLEITQNHIS 300

QY 301 AHVNSPEVGKVNEDLDIVSQSGDLTISFNPTYLIESLKAISKETVKIHFLSPVPPFTLT 360  
 DB 301 AHVNSPEVGKVNEDLDIVSQSGDLTISFNPTYLIESLKAISKETVKIHFLSPVPPFTLT 360

QY 361 PGDEESFIQLITPVRTN 378  
 DB 361 PGDEESFIQLITPVRTN 378

RESULT 2  
 ABP27970  
 ID ABP27970 standard; protein; 378 AA.  
 XX  
 AC ABP27970;  
 XX  
 DT 02-JUL-2002 (first entry)  
 XX  
 DE Streptococcus polypeptide SEQ ID NO 5116.  
 XX  
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX  
 OS Streptococcus agalactiae.  
 XX  
 PN WO200234771-A2.  
 XX  
 PD 02-MAY-2002.  
 XX  
 PF 29-OCT-2001; 2001WO-GB004789.  
 XX  
 PR 27-OCT-2000; 2000GB-00026333.  
 PR 24-NOV-2000; 2000GB-00028727.  
 PR 07-MAR-2001; 2001GB-00005640.  
 XX  
 PA (CHTR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;  
 PI Tettelin H;  
 XX  
 DR WPI; 2002-352536/38.  
 DR N-PSDB; AEN68601.  
 XX  
 PT New Streptococcus protein for the treatment or prevention of infection or  
 PT disease caused by Streptococcus bacteria, such as meningitis, and for  
 PT detecting a compound that binds to the protein.

XX PS Claim 1; Page 3675; 4525pp; English.  
 XX CC The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins  
 XX SQ Sequence 378 AA;

Query Match 99.2%; Score 1849; DB 5; Length 378;  
 Best Local Similarity 99.2%; Pred. No. 1.2e-154;  
 Matches 375; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MIQFSINRTLFHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNQISIENTIPV 60  
 DB 1 MIQFSINRTLFHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGNQISIENTIPV 60

QY 61 SNEAGLLITSPGAILLEASFFINIISLPDISINVKIEHQVLTSGKSEITLKGKDV 120  
 DB 61 SNEAGLLITSPGAILLEASFFINIISLPDISINVKIEHQVLTSGKSEITLKGKDV 120

QY 121 DQYPRLEQVSTENPLIKTKLLSIIAETAFASLOESRPILTCVHIVLSNKHDFKAVAT 180  
 DB 121 DQYPRLEQVSTENPLIKTKLLSIIAETAFASLOESRPILTCVHIVLSNKHDFKAVAT 180

QY 181 DSHRMSQRLITLNTSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240  
 DB 181 DSHRMSQRLITLNTSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240

QY 241 YTRLLEGNYPTDRLMTTEFETEVNTQSLRHAMERAFLLISNATONGTVKLEITQNHIS 300  
 DB 241 YTRLLEGNYPTDRLMTTEFETEVNTQSLRHAMERAFLLISNATONGTVKLEITQNHIS 300

QY 301 AHVNSPEVGKVNEDLDIVSQSGDLTISFNPTYLIESLKAISKETVKIHFLSPVPPFTLT 360  
 DB 301 AHVNSPEVGKVNEDLDIVSQSGDLTISFNPTYLIESLKAISKETVKIHFLSPVPPFTLT 360

QY 361 PGDEESFIQLITPVRTN 378  
 DB 361 PGDEESFIQLITPVRTN 378

RESULT 3  
 ABG61623  
 ID ABG61623 standard; protein; 378 AA.  
 XX  
 AC ABG61623;  
 XX  
 DT 12-AUG-2002 (first entry)  
 XX  
 DE Streptococcus pyogenes DnaN protein beta sub-unit.  
 XX  
 KW DNA polymerase III holoenzyme subunit protein; replicase; polymerase;  
 KW antibacterial; vaccine.  
 XX  
 OS Streptococcus pyogenes.  
 XX  
 PN WO200234936-A2.  
 XX

PD	02-MAY-2002.	AC	ABU46377;
XX		XX	
PF	29-OCT-2001; 2001WO-US048396.	DT	19-JUN-2003 (first entry)
XX		XX	
PR	27-OCT-2000; 2000US-0244023P.	DE	Protein encoded by prokaryotic essential gene #31904.
XX		XX	
PA	(REPL-) REPLIDYNE INC.	KW	Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX		XX	
PI	Mchenry CS, Bullard JM, Janjic N, Manhardt EL, Kery V;	OS	Streptococcus pyogenes.
PI	Williams JC;	PN	WO200277183-A2.
XX		XX	
DR	WPI; 2002-463366/49.	XX	
DR	N-PSDB; ABK83362, ABK83362.	PD	03-OCT-2002.
XX		XX	
PT	Novel bacterial DNA polymerase III subunit proteins and genes encoding	PF	21-MAR-2002; 2002WO-US009107.
PT	the proteins, for reconstituting replicases and polymerases for	XX	
PT	sequencing, amplification and screening compounds that modulate	PR	21-MAR-2001; 2001US-00815242.
PT	polymerase.	PR	06-SEP-2001; 2001US-00948993.
XX		PR	25-OCT-2001; 2001US-0342923P.
XX		PR	08-FEB-2002; 2002US-00072851.
PS	Claim 9; Fig 21B; 268pp; English.	PR	06-MAR-2002; 2002US-0362699P.
XX		XX	
CC	The invention describes an isolated bacterial, preferably Streptococcus	PA	(ELIT-) ELITRA PHARM INC.
CC	pyogenes DNA polymerase III holoenzyme subunit proteins (I). The protein	XX	
CC	and the polynucleotide encoding it are useful for reconstituting	PI	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
CC	replicases and polymerases for sequencing, amplification and screening	PI	Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
CC	for compounds which modulate the function of polymerase or replicase. An	XX	
CC	antibody to the protein is useful as therapeutic compound to passively	DR	WPI; 2003-029926/02.
CC	immunise an animal to protect the animal from bacteria susceptible to	XX	
CC	treatment by such antibodies, preferably Streptococcus pyogenes, as	DR	N-PSDB; ACA50247.
CC	reagents in assays to detect infection by such bacteria, and/or as tools	XX	
CC	to screen expression libraries and/or to recover desired proteins from a	PT	New antisense nucleic acids, useful for identifying proteins or screening
CC	mixture of protein and other contaminants. This is the amino acid	PT	for homologous nucleic acids required for cellular proliferation to
CC	sequence of a Streptococcus pyogenes polymerase III holoenzyme sub-unit	PT	isolate candidate molecules for rational drug discovery programs.
XX		XX	
SQ	Sequence 378 AA;	PS	Claim 25; SEQ ID NO 74301; 1766pp; English.
		XX	
Query Match	99.2%; Score 1849; DB 5; Length 378;	CC	The invention relates to an isolated nucleic acid comprising any one of
Best Local Similarity	99.2%; Pred. No. 1.2e-154;	CC	the 6213 antisense sequences given in the specification where expression
Matches 375; Conservative	1; Mismatches 2; Indels 0; Gaps 0;	CC	of the nucleic acid inhibits proliferation of a cell. Also included are:
		CC	(1) a vector comprising a promoter operably linked to the nucleic acid
QY	1 MIQFSINTLFLHALNTTKRAISTKNAIPILSSIKIEVTSTGVLTGSGNQISIENTIPV 60	CC	encoding a polypeptide whose expression is inhibited by the antisense
Db		CC	nucleic acid; (2) a host cell containing the vector; (3) an isolated
		CC	polypeptide or its fragment whose expression is inhibited by the
		CC	antisense nucleic acid; (4) an antibody capable of specifically binding
QY	61 SNEAGLLITSPGAILLEASFFINISSLPDISINVKEIQHVLTSGKSEITLKGKDV 120	CC	the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
Db		CC	proliferation or the activity of a gene in an operon required for
		CC	proliferation; (7) identifying a compound that influences the activity of
		CC	the gene product or that has an activity against a biological pathway
QY	121 DOYPRLOEVSTENPLILTKLKSIIAETAFASLOESRPILTGVHIVLSNHNKDKAVAT 180	CC	required for proliferation, or that inhibits cellular proliferation; (8)
Db		CC	identifying a gene required for cellular proliferation or the biological
		CC	pathway in which a proliferation-required gene or its gene product lies
		CC	or a gene on which the test compound that inhibits proliferation of an
QY	181 DSHRMSQRLITLNDTSADLMVLPKSLRPEPSAVFTDDIETVEVFPSPQILFRSEHLSF 240	CC	organism acts; (9) manufacturing an antibiotic; (10) profiling a
Db		CC	compound's activity; (11) a culture comprising strains in which the gene
		CC	product is overexpressed or underexpressed; (12) determining the extent
		CC	to which each of the strains is present in a culture or collection of
QY	241 YTRLLEGNYPDTDRLLMTFEFEVFNFTQSLRHAMERAFLLSNATQNGTVKLEITQNHIS 300	CC	strains; or (13) identifying the target of a compound that inhibits the
Db		CC	proliferation of an organism. The antisense nucleic acids are useful for
		CC	identifying proteins or screening for homologous nucleic acids required
		CC	for cellular proliferation to isolate candidate molecules for rational
QY	301 AHVNSPEVGNKVEDIDIVSQSGSDLTISFNPTYLIESLKAISKSETVKIHFLSPVPFPTLT 360	CC	drug discovery programs, or for screening homologous nucleic acids
Db		CC	required for proliferation in cells other than S. aureus, S. typhimurium,
		CC	K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
		CC	the target prokaryotic essential genes. Note: The sequence data for this
QY	361 PGDEESFQLITPVRTN 378	CC	patent did not form part of the printed specification, but was obtained
Db		CC	in electronic format directly from WIPO at
		CC	ftp.wipo.int/pub/published_pct_sequences
		XX	
		SQ	Sequence 378 AA;
RESULT 4			
ABU46377	Query Match		99.2%; Score 1849; DB 6; Length 378;
ID	Best Local Similarity		99.2%; Pred. No. 1.2e-154;
XX	ABU46377 standard; protein; 378 AA.		Matches 375; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MIQFSINRTLFHALNTTKRAISTKNAIPILSSIKIEVSTGVTLTSGNGQISIENTIPV 60  
 DDb 1 MIQFSINRTLFHALNTTKRAISTKNAIPILSSIKIEVSTGVTLTSGNGQISIENTIPV 60  
 QY 61 SNEAGLLITSPGAILLEASFFINIISLPPDISINVKIEHQVVLTSKSEITLKGKDV 120  
 DDb 61 SNEAGLLITSPGAILLEASFFINIISLPPDISINVKIEHQVVLTSKSEITLKGKDV 120  
 QY 121 DQYPRLOEVSTENPLILKTLKLSIIAETAFASLOESRPILTGVIHVLNKHDKFKAAT 180  
 DDb 121 DQYPRLOEVSTENPLILKTLKLSIIAETAFASLOESRPILTGVIHVLNKHDKFKAAT 180  
 QY 181 DSHRMSORLITLNTSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240  
 DDb 181 DSHRMSORLITLNTSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240  
 QY 241 YTRLLEGNYPDRLMLTETFEVNTQSLRHAMERAFILSNATQNGTVKLEITQNHIS 300  
 DDb 241 YTRLLEGNYPDRLMLTETFEVNTQSLRHAMERAFILSNATQNGTVKLEITQNHIS 300  
 QY 301 AHVNSPEVGKVNEDLDIVSQSGDLTISFNPTVILIESLKAISKETVKIHFLSPVRPFTLT 360  
 DDb 301 AHVNSPEVGKVNEDLDIVSQSGDLTISFNPTVILIESLKAISKETVKIHFLSPVRPFTLT 360  
 QY 361 PGDEESFIQLITPVRTN 378  
 DDb 361 PGDEESFIQLITPVRTN 378

RESULT 5  
 ABU44094  
 ID ABU44094 standard; protein; 378 AA.  
 AC ABU44094;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #29621.  
 XX  
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 OS Streptococcus mutans.  
 XX  
 XX WO200277183-A2.  
 XX  
 XX 03-OCT-2002.  
 XX  
 XX 21-MAR-2002; 2002WO-US009107.  
 XX  
 XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 XX (ELIT-) ELITRA PHARM INC.  
 XX  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI: 2003-029926/02.  
 DR N-PSDB; ACA47964.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 25; SEQ ID NO 72018; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 378 AA;

Query Match. 82.0%; Score 1527; DB 6; Length 378;  
 Best Local Similarity 79.9%; Pred. No. 3.7e-126;  
 Matches 302; Conservative 39; Mismatches 37; Indels 0; Gaps 0;  
 QY 1 MIQFSINRTLFHALNTTKRAISTKNAIPILSSIKIEVSTGVTLTSGNGQISIENTIPV 60  
 DDb 1 MIQFSINRTLFHALNTTKRAISTKNAIPILSSIKIEVSTGVTLTSGNGQISIENTIPV 60  
 QY 61 SNEAGLLITSPGAILLEASFFINIISLPPDISINVKIEHQVVLTSKSEITLKGKDV 120  
 DDb 61 SNEAGLLITSPGAILLEASFFINIISLPPDISINVKIEHQVVLTSKSEITLKGKDV 120  
 QY 121 DQYPRLOEVSTENPLILKTLKLSIIAETAFASLOESRPILTGVIHVLNKHDKFKAAT 180  
 DDb 121 DQYPRLOEVSTENPLILKTLKLSIIAETAFASLOESRPILTGVIHVLNKHDKFKAAT 180  
 QY 181 DSHRMSORLITLNTSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240  
 DDb 181 DSHRMSORLITLNTSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHISF 240  
 QY 241 YTRLLEGNYPDRLMLTETFEVNTQSLRHAMERAFILSNATQNGTVKLEITQNHIS 300  
 DDb 241 YTRLLEGNYPDRLMLTETFEVNTQSLRHAMERAFILSNATQNGTVKLEITQNHIS 300  
 QY 301 AHVNSPEVGKVNEDLDIVSQSGDLTISFNPTVILIESLKAISKETVKIHFLSPVRPFTLT 360  
 DDb 301 AHVNSPEVGKVNEDLDIVSQSGDLTISFNPTVILIESLKAISKETVKIHFLSPVRPFTLT 360  
 QY 361 PGDEESFIQLITPVRTN 378  
 DDb 361 PGDEESFIQLITPVRTN 378

RESULT 6  
 ABP27969  
 ID ABP27969 standard; protein; 378 AA.  
 XX ABP27969;  
 AC ABP27969;  
 XX  
 DT 02-JUL-2002 (first entry)  
 XX  
 DE Streptococcus polypeptide SEQ ID NO 5114.

XX Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.  
 XX Streptococcus pyogenes.  
 OS Streptococcus pyogenes.  
 PN WO200234771-A2.  
 XX 02-MAY-2002.  
 XX 29-OCT-2001; 2001WO-GB004789.  
 XX 27-OCT-2000; 2000GB-00026333.  
 PR 24-NOV-2000; 2000GB-00028727.  
 PR 07-MAR-2001; 2001GB-00005640.  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX Telford J, Masignani V, Margarit Y Rosl, Grandi G, Fraser C;  
 PI Tettelin H;  
 XX WPI; 2002-352536/38.  
 DR N-PSDB; ABN68600.  
 XX New Streptococcus protein for the treatment or prevention of infection or  
 PT disease caused by Streptococcus bacteria, such as meningitis, and for  
 PT detecting a compound that binds to the protein.  
 XX Claim 1; Page 3675; 4525pp; English.  
 XX The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins  
 XX Sequence 378 AA;  
 SQ  
 Query Match 80.0%; Score 1491; DB 5; Length 378;  
 Best Local Similarity 77.5%; Pred. No. 5.7e-123; Indels 0; Gaps 0;  
 Matches 293; Conservative 39; Mismatches 46;  
 QY 1 MIQFSINTFLTHALNTTAKRAISTKNAIPILSSIKIEVTSTGVTLTSGNGQISIENTIPV 60  
 DB 1 MHFSSINKFFHALVTVKRAISHKNAIPILSTVKIEVTRDAIILTSGNGQISIENTIPA 60  
 QY 61 SNNAGLLITSPGAILLGRASPPINIISSLPDISINVKIEHQHVLTSGKSEITLKGKDV 120  
 DB 61 SNNAGLLITSPGAILLGRASPPINIISSLPDISINVKIEHQHVLTSGKSEITLKGKDV 120  
 QY 121 DOYPLQVSTENPLILTKLLKSIATFAASIQESRPILTGVHVLNKHDKPKAVAT 180  
 DB 121 DOYPLQVSTENPLILTKLLKSIATFAASIQESRPILTGVHVLNKHDKPKAVAT 180  
 QY 181 DSHRMSQRLITLIDNTSLMVLVPSKLSREFSAVFTDDIETVEVFPSPQILFRSEHSIF 240  
 DB 181 DSHRMSQRTFQLEKANNFDLVVPSKLSREFSAVFTDDIETVEVFPSPQILFRSEHSIF 240  
 QY 241 YTRLLEGNYPTDRLLTNQTETFEIIFNTNALRHAMERAYLISNATQNGTVRLBQNETVS 300

DB 241 YTRLLEGNYPTDRLLTNQTETFEIIFNTNALRHAMERAYLISNATQNGTVRLBQNETVS 300  
 QY 301 AHVNSPEVGKVNEDLDIVSQSGDLTISFNFTYILIESLKAISKETVKIHFSPVRPFFLT 360  
 DB 301 AHVNSPEVGKVNEDLDIVSQSGDLTISFNFTYILIESLKAISKETVKIHFSPVRPFFLT 360  
 QY 361 PGDEESFIQLITPVRTN 378  
 DB 361 PGDETFIQLITPVRTN 378  
 RESULT 7  
 ABU02772  
 ID ABU02772 standard; protein; 378 AA.  
 XX AC ABU02772;  
 XX 23-OCT-2003 (revised)  
 DT 11-FEB-2003 (first entry)  
 XX S. pneumoniae type 4 strain protein from coding region #2352.  
 DE Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;  
 KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;  
 KW gene therapy; vaccine.  
 XX Streptococcus pneumoniae; type 4 strain.  
 OS WO200277021-A2.  
 PN 03-OCT-2002.  
 PD 27-MAR-2002; 2002WO-IB002163.  
 XX 27-MAR-2001; 2001GB-00007658.  
 PR (CHIR-) CHIRON SPA.  
 XX (GENO-) INST GENOMIC RES.  
 PI Masignani V, Tettelin H, Fraser C;  
 XX WPI; 2003-040579/03.  
 DR N-PSDB; ABX08064.  
 XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or  
 PT ear infection.  
 XX Claim 1; SEQ ID NO 4704; 56pp; English.  
 XX The invention relates to a protein comprising or having at least 50%  
 CC identity to any of the 2489 amino acid sequences, identified in the  
 CC specification (available on a computer readable format), or its fragment,  
 CC expressed from 2469 of 2489 identified DNA coding regions from the  
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
 CC AB556454. Also included are an antibody which binds one of the proteins,  
 CC treating a patient by administering the protein, DNA or antibody (in a  
 CC composition), a kit comprising first and second primers, which are the  
 CC nucleic acid cited above or fragments between nucleotides 8-100 of a  
 CC sequence not defined in the specification, for amplifying a target  
 CC the first primer is substantially complementary to the target sequence  
 CC and the second primer is substantially complementary to the complement of  
 CC the target sequence, and where the parts of the primers having  
 CC substantial complementarity define the termini of the target sequence to  
 CC be amplified, assay comprising contacting a test compound with the  
 CC protein, and determining whether the test compound binds to the protein  
 CC and a Streptococcus pneumoniae bacterium, where one or more genes  
 CC encoding the proteins has been rendered inactive. The proteins, nucleic  
 CC acid molecules, antibody and compositions are useful as medicaments for  
 CC treating or preventing a disease or infection due to streptococcus

CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
 CC media or ear infection. They are also useful in developing vaccines,  
 CC diagnostics and antibiotics. The methods are useful for identifying,  
 CC immunodominant proteins. The present sequence is one of the 2469 proteins  
 CC expressed by the identified coding regions from the genomic sequence.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
 CC standardise OS field)  
 CC  
 XX

SQ Sequence 378 AA;

Query Match 74.6%; Score 1389; DB 6; Length 378;

Best Local Similarity 72.2%; Pred. No. 6e-114;

Matches 273; Conservative 50; Mismatches 55; Indels 0; Gaps 0;

QY 1 MIOFSINRFLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTSGNGQISIENTIPV 60  
 DB 1 MIHFSINKNLFQALNTTKRAISSKNAIPILSTVKIDVTNEGITLIGSNGQISIENTISQ 60  
 QY 61 SNENAGLLITSPGAILLEASFFINIISSLPDISINVKIEIQHVLTSGSKSEITLKGDV 120  
 DB 61 KNEDAGLLITSLGSIILEASFFINNVSSLPDVLDFKEIQNVLTSGSKSEITLKGDV 120  
 QY 121 DOYPRLOEVSTENPILKTLKLSIAETAFAASIQESRPILTGVIHVLNKHOKFVAT 180  
 DB 121 EQYPRIOEISASTPLILETKLKKIINETAFAASTQESRPILTGVIHVLNKHOKFVAT 180  
 QY 181 DSHRMSORLITLDMTASDLMLVLPKSLREFSAVTDIETVEVFPSPQILFREHSIF 240  
 DB 181 DSHRLSOKKLTLEKNSSDDFDVIPSRLREFSAVTDIETVEIFFANQLFRSENISF 240  
 QY 241 YTRLLEGNYPTDRLMTFETEVVFNQSLRHAMERAFILSNATQNGTVKLEITQNHIS 300  
 DB 241 YTRLLEGNYPTDRLIPTDFNTITFNVNLRQSMERARLLSSATQNGTVKLEIKDGVV 300  
 QY 301 AHVNSPEVGVKNEDIDIVSQSGDLTISFNPTLIESKATKSETVKIHFSPVRPFTLT 360  
 DB 301 AHVHSPGVKNEDIDTQVGTGDLTISFNPTLIDSLKALNSEKVTISFISAVRPFTLV 360  
 QY 361 PGDEESFIQLITPVRTN 378  
 DB 361 PADTDEDFMQLITPVRTN 378

RESULT 8

ADK47890

ID ADK47890 standard; protein; 378 AA.

XX

AC ADK47890;

XX 20-MAY-2004 (first entry)

DT Streptococcus pneumoniae protein, Seq ID No 4405.

XX Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.

XX Streptococcus pneumoniae.

XX US6699703-B1.

XX 02-MAR-2004.

XX 26-MAY-2000; 2000US-005693110.

XX 02-JUL-1997; 97US-0051553P.

XX 12-MAY-1998; 98US-0085131P.

XX 30-JUN-1998; 98US-00107433.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;

XX

DR WPI; 2004-2123399/20.

DR N-PSDB; ADK45229.

XX New nucleic acid molecules and polypeptides useful for diagnosing,  
 PT preventing and treating pathological conditions resulting from bacterial  
 PT infection, e.g. Streptococcus pneumoniae infection, and in drug  
 PT screening.

PS Disclosure; SEQ ID NO 4405; 301pp; English.

XX The invention relates to isolated Streptococcus pneumoniae nucleic acids  
 CC and polypeptides. The nucleic acids and proteins are useful for  
 CC diagnosing, preventing and treating pathological conditions resulting  
 CC from bacterial infection, such as S. pneumoniae infection. These may also  
 CC be used for drug screening procedures. The present sequence represents a  
 CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence  
 CC data for this patent did not appear in the printed specification but was  
 CC obtained in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html.

SQ Sequence 378 AA;

Query Match 74.6%; Score 1389; DB 8; Length 378;

Best Local Similarity 72.2%; Pred. No. 6e-114;

Matches 273; Conservative 50; Mismatches 55; Indels 0; Gaps 0;

QY 1 MIOFSINRFLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTSGNGQISIENTIPV 60  
 DB 1 MIHFSINKNLFQALNTTKRAISSKNAIPILSTVKIDVTNEGITLIGSNGQISIENTISQ 60  
 QY 61 SNENAGLLITSPGAILLEASFFINIISSLPDISINVKIEIQHVLTSGSKSEITLKGDV 120  
 DB 61 KNEDAGLLITSLGSIILEASFFINNVSSLPDVLDFKEIQNVLTSGSKSEITLKGDV 120  
 QY 121 DOYPRLOEVSTENPILKTLKLSIAETAFAASIQESRPILTGVIHVLNKHOKFVAT 180  
 DB 121 EQYPRIOEISASTPLILETKLKKIINETAFAASTQESRPILTGVIHVLNKHOKFVAT 180  
 QY 181 DSHRMSORLITLDMTASDLMLVLPKSLREFSAVTDIETVEVFPSPQILFREHSIF 240  
 DB 181 DSHRLSOKKLTLEKNSSDDFDVIPSRLREFSAVTDIETVEIFFANQLFRSENISF 240  
 QY 241 YTRLLEGNYPTDRLMTFETEVVFNQSLRHAMERAFILSNATQNGTVKLEITQNHIS 300  
 DB 241 YTRLLEGNYPTDRLIPTDFNTITFNVNLRQSMERARLLSSATQNGTVKLEIKDGVV 300  
 QY 301 AHVNSPEVGVKNEDIDIVSQSGDLTISFNPTLIESKATKSETVKIHFSPVRPFTLT 360  
 DB 301 AHVHSPGVKNEDIDTQVGTGDLTISFNPTLIDSLKALNSEKVTISFISAVRPFTLV 360  
 QY 361 PGDEESFIQLITPVRTN 378  
 DB 361 PADTDEDFMQLITPVRTN 378

RESULT 9

AAU37853

ID AAU37853 standard; protein; 378 AA.

XX

AC AAU37853;

XX 14-FEB-2002 (first entry)

DT Streptococcus pneumoniae cellular proliferation protein #282.

XX Antisense; prokaryotic cellular proliferation protein; antibiotic;

XX antibacterial; drug design.

XX Streptococcus pneumoniae.

XX WO200170955-A2.

XX 27-SEP-2001.

```
XX 21-MAR-2001; 2001WO-US009180.
XX 21-MAR-2000; 2000US-0191078P.
XX 23-MAY-2000; 2000US-0206848P.
XX 26-MAY-2000; 2000US-0207272P.
XX 23-OCT-2000; 2000US-0242578P.
XX 27-NOV-2000; 2000US-0253625P.
XX 22-DEC-2000; 2000US-0257931P.
XX 16-FEB-2001; 2001US-0269308P.
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlseen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
XX N-PSDB; NAS55712.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids.
XX
XX Example 3; SEQ ID NO 13446; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the genes,
XX their use in the discovery of novel antibiotics, the essential genes,
XX themselves and the encoded proteins. The prokaryotes used are Escherichia
XX coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
XX Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
XX useful for the identification of potential new targets for antibiotic
XX development. The antisense nucleic acids can also be used to identify
XX proteins used in proliferation, to express these proteins, and to obtain
XX antibodies capable of binding to the expressed proteins. The proteins can
XX be used to screen compounds in rational drug discovery programmes. The
XX antisense nucleic acid sequence is also useful to screen for homologous
XX nucleic acids which are required for cell proliferation in a wide variety
XX of organisms. The present sequence represents an essential prokaryotic
XX cellular proliferation protein. Note: The sequence data for this patent
XX did not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 378 AA;
XX
XX Query Match 74.3%; Score 1384; DB 4; Length 378;
XX Best Local Similarity 72.2%; Pred. No. 1.6e-113;
XX Matches 273; Conservative 49; Mismatches 56; Indels 0; Gaps 0;
XX
XX 1 MIQFSINTLFIHALNTYKRAISTKNAIPILSSIKIEVTSTGVTTGNGQISIENTIPV 60
XX 1 MIHFSINKNLFQALNITKRAISSKNAIPILSTKIDVTNEGVTLIGSNGQISIENTISQ 60
XX
XX 61 SNEAGLLITSPGATLLAASFPINISLIPDISINVKETEOHQVVLTSKSTITLKGKDV 120
XX 61 KNEAGLLITSGSLLAASFPINVSLLPDLDFKEQIVTLTSKSTITLKGKDS 120
XX
XX 121 DQYPRLOEVSTENPLILTKLKLSIAETAFAASLQESRPILTVGHVILSNHKKFVAVT 180
XX 121 EQYPRIQEISASTPLILETKLKLIINETAFAASTQESRPILTVGHVILSQHKLKTAVT 180
XX
XX 181 DSHRMSQRILTDNTSADLMVLPKSLUREFSAVTFDDIETVEVFPSPQILFRSEHISF 240
XX 181 DSHRLSKQKLTLEKNSDDPDVVPISRLREFSAVTFDDIETVEIFPANNQILFRSENISF 240
XX
XX 241 YTRLLEGNYPDTRLIMLTETEFETEVNTOSTLRHMERFLISNATQNGVTKLEITQNHIS 300
XX 241 YTRLLEGNYPDTRLIPTDFTNTITFNVNVLQSMERARLLSSATQNGVTKLEIKDGVVS 300
XX
XX 301 AHVNSPEVGNVEDLDIVSQSGDLTISFNPLYIESKAIKSETVKIHFSLPVPFFTLT 360
XX 301 AHVHSPVGVKNNEEDTDQVGTEDLTISFNPLYLIDSLKALNSEKVTISFISAVPFTLV 360
```

```
QY 361 PGDEERSFIQLITPVRTN 378
DB 361 PADTDEDFMQLITPVRTN 378
RESULT 10
ABU45758
ID ABU45758 standard; protein; 378 AA.
XX AC ABU45758;
XX DT 19-JUN-2003 (first entry)
XX DE Protein encoded by Prokaryotic essential gene #31285.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Streptococcus pneumoniae.
XX PN WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlseen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX N-PSDB; ACA49628.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 73682; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX proliferation of an organism. The antisense nucleic acids are useful for
XX identifying proteins or screening for homologous nucleic acids required
XX for cellular proliferation to isolate candidate molecules for rational
XX drug discovery programs, or for screening homologous nucleic acids
XX required for proliferation in cells other than S. aureus, S. typhimurium,
XX K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX the target prokaryotic essential genes. Note: The sequence data for this
```



CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 378 AA;

Query Match 74.3%; Score 1384; DB 6; Length 378;  
 Best Local Similarity 72.2%; Pred. No. 1.6e-113;  
 Matches 273; Conservative 49; Mismatches 56; Indels 0; Gaps 0;

QY 1 MIOFSINRTLFHAIANTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGQISIENTIPV 60  
 DB 1 MIHFSINKNLFQALNITKRAISSKNAIPILSVKIDVTNEGVTLTGSGQISIENTISQ 60  
 QY 61 SNEAGLITSPGAILLEASFFINIISLPDISINVKIEQHGVLTSGKSEITLKGKDV 120  
 DB 61 KNEAGLITSLGILLESFFINNVSSLPDVTLDKFEIQVLTSGKSEITLKGKDS 120  
 QY 121 DQYPRLOEVSTENPLILKTLKLSIIAETAFAASLOESRPILTGVIHVLNSHKDFKAVAT 180  
 DB 121 EQYPRIOEISASPLILETKLKKIINETAFAASTQESRPILTGVIHVLNSHKELKTVAT 180  
 QY 181 DSHRMSORLITDNTSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHSIF 240  
 DB 181 DSHRLSQKLTLEKNSDDFDVIPSRLREFSAVFTDDIETVEIIPFANNQILFRSENISF 240  
 QY 241 YTRLLEGNYPDTRLMTTEFETEVVFNTOQLRHAMERAPLISNATQNGTVKLEITQNHIS 300  
 DB 241 YTRLLEGNYPDTRLIPTDFNTIIFNVNLRQSMERARLLSSATQNGTVKLEIKDGVVS 300  
 QY 301 AHVNSPEVGNVEDLDIVSQSGDLTISFNPTYLIESLKAISKETVKIHFSPVPRPFTLT 360  
 DB 301 AHVNSPEVGNVEEDTDQVTGEDLTISFNPTYLIDSLKALNSEKVTISFISAVRPFTLV 360  
 QY 361 PGDEESFQILITPVRTN 378  
 DB 361 PADTDEDFQILITPVRTN 378

RESULT 11

ABBS3301  
 ID ABBS3301 standard; protein; 380 AA.

AC ABBS3301;

XX 29-AUG-2003 (revised)  
 DT 16-MAY-2002 (first entry)

XX Lactococcus lactis protein dnan.

XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.

XX Lactococcus lactis; IL1403.

XX FR2807446-A1.

XX 12-OCT-2001.

XX 11-APR-2000; 2000FR-00004630.

XX 11-APR-2000; 2000FR-00004630.

XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.

XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;

XX WPI; 2002-043418/06.

XX New nucleotide sequence useful in the identification or Lactococcus

XX lactic acid and related species.

XX Claim 6; SEQ ID NO 3; 2504pp; French.

CC The present invention is related to a Lactococcus lactis nucleotide  
 CC sequence (ABA90521) and related proteins (ABBS3300-ABBS5621). The nucleic  
 CC acid sequence is useful in the detection and/or amplification of nucleic  
 CC acid sequence, particularly to identify Lactococcus lactis or related  
 CC species. The proteins of the invention are useful for the biosynthesis or  
 CC biodegradation of a composition of interest. The invention helps research  
 CC in lactic bacteria, particularly useful in the production of yogurt and  
 CC cheese. Note: The sequence data for this patent is based on equivalent  
 CC patent WO2001/77334 (published 18-OCT-2001) which is available in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 29-AUG-2003 to  
 CC standardise OS field)

XX Sequence 380 AA;

Query Match 56.5%; Score 1053; DB 5; Length 380;  
 Best Local Similarity 54.7%; Pred. No. 3.1e-84;  
 Matches 208; Conservative 73; Mismatches 97; Indels 2; Gaps 1;

QY 1 MIOFSINRTLFHAIANTTKRAISTKNAIPILSSIKIEVTSTGVTLTGSGQISIENTIPV 60  
 DB 1 MIKFSINKNAFQNALRITKQAIGSKVTIPALTKLIEVEENGITLTGSGQISIKNPLP 60  
 QY 61 SNEAGLITSPGAILLEASFFINIISLPDISINVKIEQHGVLTSGKSEITLKGKDV 120  
 DB 61 DNKASMLISGTGSLLEAAFFENVVSQLPEVLTETKQKQVLLTSGKSEITLKGDS 120  
 QY 121 DQYPRLOEVSTENPLILKTLKLSIIAETAFAASLOESRPILTGVIHVLNSHKDFKAVAT 180  
 DB 121 EYPHLOEISEGSLKMKVKVLEIFTEITVFANSTQENRPIFTGVHLETLSTGELKAVAT 180  
 QY 181 DSHRMSORLITDNTSADLMVLPKSLREFSAVFTDDIETVEVFPSPQILFRSEHSIF 240  
 DB 181 DSHRMSORLITPLESELKFDVILPSKINSFKVFTNDEEIEIFISGQILFRNETISY 240  
 QY 241 YTRLLEGNYPDTRLMTTE--FETEVVFNTOQLRHAMERAPLISNATQNGTVKLEITQNH 298  
 DB 241 YSRLIEGSDPTNRLIPNEADYTLDLVFDAAQLRHTMDRALLTVMTTNGTVKLTSGDS 300  
 QY 299 ISAHVNSPEVGNVEDLDIVSQSGDLTISFNPTYLIESLKAISKETVKIHFSPVPRPPT 358  
 DB 301 VVTTANSPEVGSVHEELTALSKEGNDLAIISFNPEYLDALKVKAPEVRIFISNVRPPT 360  
 QY 359 LTPGDEESFQILITPVRTN 378  
 DB 361 LQPRNEESGFQILITPVRTN 380

RESULT 12

AAU35314  
 ID AAU35314 standard; protein; 376 AA.

XX AAU35314;

XX 14-FEB-2002 (first entry)

XX Enterococcus faecalis cellular proliferation protein #601.

XX Antisense; prokaryotic cellular proliferation protein; antibiotic;

XX antibacterial; drug design.

XX Enterococcus faecalis.

XX WO2001/0955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US009180.

XX 21-MAR-2000; 2000US-0191078P.

XX 23-MAY-2000; 2000US-0206848P.

XX 26-MAY-2000; 2000US-0207727P.

XX 23-OCT-2000; 2000US-0242578P.

PR 27-NOV-2000; 2000US-0253625P.  
 PR 22-DEC-2000; 2000US-0257931P.  
 PR 16-FEB-2001; 2001US-0269308P.  
 XX  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX  
 DR WPI; 2001-611495/70.  
 DR N-PSDB; AAS53173.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids.  
 XX  
 XX Example 3; SEQ ID NO 10907; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the genes,  
 CC their use in the discovery of novel antibiotics, the essential genes,  
 CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
 CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
 CC pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
 CC useful for the identification of potential new targets for antibiotic  
 CC development. The antisense nucleic acids can also be used to identify  
 CC proteins used in proliferation, to express these proteins, and to obtain  
 CC antibodies capable of binding to the expressed proteins. The proteins can  
 CC be used to screen compounds in rational drug discovery programmes. The  
 CC antisense nucleic acid sequence is also useful to screen for homologous  
 CC nucleic acids which are required for cell proliferation in a wide variety  
 CC of organisms. The present sequence represents an essential prokaryotic  
 CC cellular proliferation protein. Note: The sequence data for this patent  
 CC did not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 376 AA;

ID ABU29060 standard; protein; 376 AA.  
 AC ABU29060;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #14587.  
 XX  
 XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 OS Enterococcus faecalis.  
 XX  
 PN WO200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR N-PSDB; ACA32930.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 25; SEQ ID NO 56984; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than S. aureus, S. typhimurium,  
 CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 376 AA;

RESULT 13

ABU29060

Query Match

46.4%; Score 864.5; DB 6; Length 376;

	Best Local Similarity	45.9%; Pred. No.	1.5e-67;	Mismatches	77; Conservative	77; Mismatches	126; Indels	1; Gaps	1;
QY	2 IQFSINRTLFIHALNTTKRAISTKNAIPILSSIKIEVTSQTGVLITSGNGOISIENTIPTVS	61	: : :	:	:	:	:	:	:
Dd	1 MKLTVKRSVFLOELQTVQRATSSKTPIILTGVKLIVSEDLGLTGSNADISIETFLSKD	60	: : :	:	:	:	:	:	:
QY	62 NENAGLLITSPGAILLLEASFFINISSLDPDISINVKEIQHQLVTLTGKGKDVP	121	: : :	:	:	:	:	:	:
Dd	61 DEKAQMTERIGSVIQQSFGEIRKLPEDMFTMEVDNNQAITSCKADFVNGLDAD	120	: : :	:	:	:	:	:	:
QY	122 QYRPLQEVSTENPILTKLKSIITAETAFAASLOSERPILGTGHVLSNHDKFKAVTD	181	: : :	:	:	:	:	:	:
Dd	121 NYPHLPVIDTQNOMKLPHVHLTISETGFVAWSHERSRPILGTGHFLENQK-DLVATD	179	: : :	:	:	:	:	:	:
QY	182 SHRMQSRLITLDNTSADMVLPVKSREFAVFDIDLETVEFPSPQILFRSEHSFY	241	: : :	:	:	:	:	:	:
Dd	180 SHRLESQVRPTQAEVEDNVIPGSLTELERSLTNEEWEVISIMENVLFKETMYFY	239	: : :	:	:	:	:	:	:
QY	242 TLLLEGNYPDTRLMLTFEFTEVVFNTOQLRHAMERAPLAISNATONGTVKLEITONHISA	301	: : :	:	:	:	:	:	:
Dd	240 SRLLEGNYPDTRNLRIPTSHNQIFEPVELLSAITERASSILLSHEGRNNIVLRISPDSWLV	299	: : :	:	:	:	:	:	:
QY	302 HYNSEPVGKNVEDDIVSQSGSDLTISENPYLIESLKAISKETVKTHFLSPRPPTLTP	361	: : :	:	:	:	:	:	:
Dd	300 YGNPSPEIGKVESALYNYSVEALDISFPNDYMCDALRAFDMNITVKFLSPIRPPTLEP	359	: : :	:	:	:	:	:	:
QY	362 GDDEESSFIOLITPPVRTN	378	: : :	:	:	:	:	:	:
Dd	360 TETELEDFOQLITPVRTN	376	: : :	:	:	:	:	:	:
 RESULT 14 ADC94929									
ID	ADC94929 standard; protein; 384 AA.								
XX	AC ADC94929;								
XX	DT 01-JAN-2004 (first entry)								
DE	E. faecium protein sequence SEQ ID 4556.								
XX	Vaccine; urinary tract infection; bacteraemia; endocarditis; wound; abdominal-pelvic infection.								
KW	Xenterococcus faecium.								
OS	US6583275-B1.								
PX	24-JUN-2003.								
PN	30-JUN-1998; 98US-00107532.								
PF	02-JUL-1997; 97US-0051571P.								
PR	14-MAY-1998; 98US-0085598P.								
PA	(GENO-) GENOME THERAPEUTICS CORP.								
PI	Doucette-Stamm LA, Bush D;								
DR	WIPI; 2003-799836/75.								
N-PADB;	ADC91275.								
PT	New isolated nucleic acid derived from Enterococcus faecium encoding an Enterococcus faecium polypeptide useful for detection, prevention and treatment of a pathological condition resulting from a bacterial infection.								
Example 1;	SEQ ID NO 4556; 243pp; English.  The invention relates to an isolated nucleic acid derived from Enterococcus faecium encoding an Enterococcus faecium polypeptide having one of 10 fully defined sequences given in the (or comprising 40								

```
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX N-PSDB; ACA33953.
DR
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 58007; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC
```

```

DB 240 SRLEGNYPDNRLIPSSFNTEVFSPFLAAIERASLLSHEGNNIVRISIRPDVVL 299
QY 302 HVNSPEVGKVNEDDIVSQSDLTISFNPYLTIESLKAISKSETVKIHFISPVVRPFTTTP 361
DB 300 YGNSPEIGKVEESLSYTSASSGDLIDISFNPYMKALRAFGDMSIKVKFISAIRPFTLEP 359
QY 362 GDEESFIQLITPVRTN 378
DB 360 TEDGVQFIQLITPVRTN 376
XX
RESULT 16
ASU42909
ID ABU42909 standard; protein; 377 AA.
XX
AC ABU42909;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #28436.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Staphylococcus epidermidis.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
XX N-PSDB; ACA46779.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 70833; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the
XX
```



XX 25-NOV-2002; 2002WO-US036123.  
 XX 29-NOV-2001; 2001US-0333777P.  
 PR 18-NOV-2002; 2002US-0426742P.  
 XX (AMHP ) WYETH HOLDINGS CORP.  
 XX Fletcher LD, Mcmichael JC, Russell DP, Zagursky RJ;  
 XX WPI; 2003-505284/47.  
 DR N-PSDB; ADB07183.  
 XX New Alloiooccus otitidis polynucleotides and polypeptides, useful for  
 PT treating and diagnosing diseases, drug screening assays and monitoring of  
 PT effects during drug clinical trials.  
 XX Claim 33; SEQ ID NO 1124; 1019pp; English.  
 XX The present invention describes an isolated polynucleotide (I) of  
 CC Alloiooccus otitidis genomic DNA, which encodes an antigenic protein.  
 CC Alloiooccus otitidis is a Gram-positive bacterium. Also described: (1)  
 CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an  
 CC expression vector comprising the novel isolated polynucleotide (I); its  
 CC complement, degenerate variant or fragment; (3) a genetically engineered  
 CC host cell, transfected, transformed or infected with the vector of (2);  
 CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic  
 CC composition comprising the polypeptide, its complement, biological  
 CC equivalent or fragment, or the polynucleotide that is comprised in the  
 CC expression vector; (6) a pharmaceutical composition comprising the  
 CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array  
 CC of the polypeptides of (1), their biological equivalent or fragment; (8)  
 CC immunizing against Alloiooccus otitidis by administering to a host the  
 CC immunogenic composition; (9) detecting and/or identifying Alloiooccus  
 CC otitidis in the biological sample; (10) a kit comprising a container  
 CC containing the novel polynucleotide, its degenerate variant or fragment,  
 CC or the antibody of (4); and (11) producing a polypeptide by culturing the  
 CC genetically engineered host cell under conditions suitable to produce the  
 CC polypeptide from the culture. (I) can be used in gene therapy. The  
 CC polynucleotides, polypeptides, antibodies and compositions of the present  
 CC invention can be used for treating and diagnosing diseases, drug  
 CC screening assays and monitoring of effects during drug clinical trials.  
 CC The polynucleotides are useful for expressing and detecting Alloiooccus  
 CC otitidis. The present sequence represents an Alloiooccus otitidis  
 CC antigen protein from the present invention.  
 XX SQ Sequence 400 AA;  
 Query Match 40.0%; Score 745; DB 6; Length 400;  
 Best Local Similarity 42.3%; Pred. No. 6e-57;  
 Matches 160; Conservative 68; Mismatches 146; Indels 4; Gaps 4;  
 QY 2 1QFSINRFLTHALNTTKRAISTKNAIPILSSIKIEVTSSTGVTLTGSGNQISIENTIPVS 61  
 DB 24 LKFSVKRTFEFLKVKQVIAVSSKSTIALITGKLEADNQGLTGTGSDSDISVESYLSVT 83  
 QY 62 NENAGLLITSPGAILLE-ASFFINIISSLPDISINVKIEHQVVLTSCKSEITLKGKDV 120  
 DB 84 DEGADLVIDEFGQIVLPARLFANIVQKLPDTHFKVNVSSQGOQTQITSASASFTINGIDA 143  
 QY 121 DOYPLQEVSTENPLILTKLLKLSIAETAFAASLOESRPILTGVHIVLSNKHDPKAVAT 180  
 DB 144 MSYPLHPIDLEESFTLPVDLFPKMINQTVIAVSNHESRPILTGVNLSLKEGR-LKAVAT 202  
 QY 181 DSHRMSQRLITDNT-SADLMVVLPSKSLRPFSAVFTDDIETVEFFSPQILFRSEHIS 239  
 DB 203 DSHRLSQRIQLESAPDISFDIVIGKSLTETLKLMDAD-EVRVAISDNQILFALSQQ 261  
 QY 240 FYTRLEGNYPDTRLLMTEFFEVFNQTSRSHAMERAFILSNATQNGTVKLEITQNH 299  
 DB 262 FYSLLEGKYPTDRLIPGDTTEITLDTKEIQGAVDRASLLSHEGKNNVQLTVTAEL 321  
 QY 300 SAHVNSPEVGKVNEDLDIVSQSGSDLTISFNPTYLIESLKAIKSETVKIHLFSPVRPTL 359

DB 322 BIEQSAEVGHVQBEIDFGHFQGDLLISFNPDLKALATFGGGEIKLVSTLPPFVI 381  
 QY 360 TPGDEEESFIQLITPVRT 377  
 DB 382 VPSEDQGDIFIQLITPIRT 399  
 RESULT 19  
 ADJ27077  
 ID ADJ27077 standard; protein; 400 AA.  
 XX  
 AC ADJ27077;  
 XX  
 DT 20-MAY-2004 (first entry)  
 XX  
 DE Alloiooccus otitidis protein, a novel antibacterial target SeqID 22.  
 XX growth and survival; otitis media with effusion; OME;  
 KW bacterial infection; antibacterial; antineurotic.  
 KW  
 OS Alloiooccus otitis.  
 XX  
 PN WO2003104391-A2.  
 XX  
 PD 18-DEC-2003.  
 XX  
 PF 25-NOV-2002; 2002WO-US036122.  
 XX  
 PR 29-NOV-2001; 2001US-0333777P.  
 PR 18-NOV-2002; 2002US-0426742P.  
 PR 18-NOV-2002; 2002US-0427367P.  
 XX  
 PA (AMHP ) WYETH HOLDINGS CORP.  
 PI Murphy E, Projan SJ;  
 XX WPI; 2004-053616/05.  
 DR N-PSDB; ADJ27076.  
 XX  
 PT Novel isolated Alloiooccus otitidis polypeptides useful for identifying  
 PT compounds that inhibit the activity or expression of the polypeptide and  
 PT thus are useful for treating infections caused by Alloiooccus otitidis.  
 XX  
 PS Claim 7; SEQ ID NO 22; 433pp; English.  
 XX  
 CC This invention relates to novel isolated Alloiooccus otitidis (A.  
 CC otitidis) nucleic acid molecules and encoded proteins thereof.  
 CC Specifically, it refers to proteins that are essential for the growth and  
 CC survival of the gram-positive bacterium A. otitidis, and hence provide  
 CC novel antibacterial targets. The present invention describes  
 CC pharmaceutical compositions and antisense compounds that are useful for  
 CC inhibiting activity or expression of these proteins. Furthermore, it  
 CC provides diagnostics and therapeutics that can be used to ameliorate  
 CC diseases that are associated with A. otitidis, such as otitis media with  
 CC effusion (OME) and various bacterial infections. Accordingly, these  
 CC compositions exhibit both antibacterial and antineurotic activities.  
 CC This polypeptide is an A. otitidis protein sequence of the invention.  
 XX  
 SQ Sequence 400 AA;  
 Query Match 40.0%; Score 745; DB 8; Length 400;  
 Best Local Similarity 42.3%; Pred. No. 6e-57;  
 Matches 160; Conservative 68; Mismatches 146; Indels 4; Gaps 4;  
 QY 2 1QFSINRFLTHALNTTKRAISTKNAIPILSSIKIEVTSSTGVTLTGSGNQISIENTIPVS 61  
 DB 24 LKFSVKRTFEFLKVKQVIAVSSKSTIALITGKLEADNQGLTGTGSDSDISVESYLSVT 83  
 QY 62 NENAGLLITSPGAILLE-ASFFINIISSLPDISINVKIEHQVVLTSCKSEITLKGKDV 120  
 DB 84 DEGADLVIDEFGQIVLPARLFANIVQKLPDTHFKVNVSSQGOQTQITSASASFTINGIDA 143





PD 29-JUL-1999.  
 XX 25-JAN-1999; 99WO-US001547.  
 PF 27-JAN-1998; 98US-0074522P.  
 PR 22-JUL-1998; 98US-0093727P.  
 XX (UYRQ) UNIV ROCKEFELLER.  
 PA O'donnell ME, Zhang D, Whipple R;  
 PI WPI; 1999-590685/50.  
 XX N-PSDB; AA231005.  
 XX New isolated dnaE, dnaX and dnaB genes from Gram positive bacteria, used  
 PT to develop screening assays for identifying antibiotic compounds.  
 XX Disclosure; Page 30-32; 132pp; English.  
 PS This is the amino acid sequence encoded by the Staphylococcus aureus  
 CC partial dnaN gene. The invention relates to a number of isolated DNA  
 CC molecules from Gram positive bacterium, corresponding to dnaE (AA231001),  
 CC dnaX (AA231002), and dnaB (AA231003). The PolC, dnaN and dnaG genes  
 CC (AA231004-231006) are also identified. The dnaB gene corresponds to the  
 CC alpha subunit of the Escherichia coli, DNA polymerase III holoenzyme,  
 CC dnaX corresponds to the gamma and tau subunits, and dnaB corresponds to  
 CC the helicase. The alpha subunit is the actual DNA polymerase, the gamma  
 CC complex forms the clamp loader and tau is a "glue protein". dnaX encodes  
 CC both gamma and tau, Tau is the product of the full gene, while gamma is  
 CC the product of the first two thirds of the gene. dnaN forms the beta  
 CC subunit which forms the sliding clamp, and dnaG encodes a primase. The  
 CC DNA sequences of the invention can be used to identify agents that  
 CC inhibit or promote DNA replication by acting on various parts of the gram  
 CC positive bacterial DNA polymerase holoenzyme. The products and methods of  
 CC the invention can be used for identifying pharmacological agents or lead  
 CC compounds for agents active at the level of a replication protein  
 CC function, particularly DNA replication. The agents identified can be used  
 CC as antibiotics  
 XX  
 SQ Sequence 377 AA;  
 Query Match 39.2%; Score 730.5; DB 2; Length 377;  
 Best Local Similarity 39.3%; Pred. No. 1e-55;  
 Matches 149; Conservative 90; Mismatches 135; Indels 5; Gaps 4;  
 QY 1 MIOFSINTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGTGTLTSGNGQISIENTIPV 60  
 Db 1 MMEFTIKRDYFTQLNDTLKALSPRTTLPILTGKIDAKEHEVILTSGDSEISITIPK 60  
 QY 61 SNEAGLL-ITSPGAILLEASFFINISSLDPISINVKIEHQHVLTSGKSEITLKGKD 119  
 Db 61 TVDGEDIVNISSETGSLVLPGRFVDIIKKLPKDKVLTSTNEQFQTLITSGHSEFNLGLD 120  
 QY 120 VDQYPRLOVSTENPILKTLKLSIAETAASLQESRPILTGTHVILSNHDKPKAVA 179  
 Db 121 PDQYILFQVSDDDAIQISVRLKVNVIATNFVSTETRPVLTGVNMLIQEN-ELICTA 179  
 QY 180 TDSHRMSORLIPLDNTSADLMVLPKSLRSEFSAVFTDDIETVEFFSPQLPSEHIS 239  
 Db 180 TDSHRLAVKQLQEDVSENKNVILPCKALAEINKIMSDNEEDIDIFFASNQVLFKGVNV 239  
 QY 240 FYTRLEGNYPDRLMTEFTTEVVVFNQSRHAMEAFILSNATONGTVKLEITQNH 299  
 Db 240 FISRLLEGHPDTRFLFPENYEIKLSIDNGEFYHAIDRASLLAREGGNNVILKLTGDDV 299  
 QY 300 SAHVNSPEVKGNEDLDIVSQSGDLTISFNPTYLIESLKATKSETVKIHFLSPVRPTL 359  
 Db 300 ELUSTSPETIGTVEEVANDVEGGSLKISFNKMMMDALKADINDVEVEFFGTMKPFIL 359  
 QY 360 TP-GDEESFIQLITPVRT 377  
 Db 360 KPKGD--DSVTQLIPIRT 376

RESULT 22  
 AAU37133  
 ID AAU37133 standard; protein; 377 AA.  
 XX  
 AC AAU37133;  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE Staphylococcus aureus cellular proliferation protein #1303.  
 KW Antisense; prokaryotic cellular proliferation protein; antibiotic;  
 KW antibacterial; drug design.  
 XX  
 OS Staphylococcus aureus.  
 XX  
 PN WO200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-US009180.  
 XX  
 PR 21-MAR-2000; 2000US-0191078P.  
 PR 23-MAY-2000; 2000US-0206848P.  
 PR 26-MAY-2000; 2000US-0207727P.  
 PR 23-OCT-2000; 2000US-0242578P.  
 PR 27-NOV-2000; 2000US-0253625P.  
 PR 22-DEC-2000; 2000US-0257931P.  
 PR 16-FEB-2001; 2001US-0269308P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX WPI; 2001-611495/70.  
 DR N-PSDB; AAS54992.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids.  
 XX  
 XX Example 3; SEQ ID NO 12726; 511pp; English.  
 PS The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the genes,  
 CC their use in the discovery of novel antibiotics, the essential genes  
 CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
 CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
 CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
 CC useful for the identification of potential new targets for antibiotic  
 CC development. The antisense nucleic acids can also be used to identify  
 CC proteins used in proliferation, to express these proteins, and to obtain  
 CC antibodies capable of binding to the expressed proteins. The proteins can  
 CC be used to screen compounds in rational drug discovery programmes. The  
 CC antisense nucleic acid sequence is also useful to screen for homologous  
 CC nucleic acids which are required for cell proliferation in a wide variety  
 CC of organisms. The present sequence represents an essential prokaryotic  
 CC cellular proliferation protein. Note: The sequence data for this patent  
 CC did not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 377 AA;  
 Query Match 39.2%; Score 730.5; DB 4; Length 377;  
 Best Local Similarity 39.3%; Pred. No. 1e-55;  
 Matches 149; Conservative 90; Mismatches 135; Indels 5; Gaps 4;  
 QY 1 MIOFSINTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGTGTLTSGNGQISIENTIPV 60  
 Db 1 MMEFTIKRDYFTQLNDTLKALSPRTTLPILTGKIDAKEHEVILTSGDSEISITIPK 60  
 QY 61 SNEAGLL-ITSPGAILLEASFFINISSLDPISINVKIEHQHVLTSGKSEITLKGKD 119



PI O'donnell ME, Bruck I, Zhang D, Whipple R;  
DR WPI; 2001-147453/15.  
DR N-PSDB; AAF54735.  
XX Isolated DNA molecule from a Gram positive bacterium encoding DNA  
PT replication proteins used to identify compounds which have antibiotic  
PT activity.  
XX  
PS Disclosure; Page 40-41; 239pp; English.  
XX  
CC The present sequence represents a dnaN polypeptide. The specification  
CC describes DNA molecules from Gram positive bacteria, which comprise a  
CC coding region from a polC, dnaB, hoiA, hoiB, dnaX, ssb, dnaG or a  
CC dnaB gene. These sequences encode proteins that replicate the chromosome  
CC of Gram positive bacteria. They are used for sequencing and amplification  
CC of DNA and in drug discovery to identify compounds which have antibiotic  
CC activity through interference with replication. They are used in methods  
CC for identifying compounds that are active at the level of DNA replication  
CC and result in arrest of cell growth or cell death of bacteria to treat  
CC bacterial infections in animals  
XX  
SQ Sequence 377 AA;

Query Match 39.2%; Score 730.5; DB 4; Length 377;  
Best Local Similarity 39.3%; Pred. No. 1e-55;  
Matches 149; Conservative 90; Mismatches 135; Indels 5; Gaps 4;  
QY 1 MIOFSINTFLTHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTSGNGOISIENTIPV 60  
DB 1 MMEFTIKRDYFITQLNDTLKAISSPTTLITGKIDAKEHEVILTGSDSISIEITIPK 60  
QY 61 SNEAGLL-ITSPGAILLEASFFINISSLDPDISINVKIEIQHQVVLTSKSEITLKGK 119  
DB 61 TVDGEDIVNISSETGSVLPGRFFVDIIKKLPGKOVKLTNEQFQTLITSGHSEFNLSGLD 120  
QY 120 VDQYPRQEVSTENPLIKTLKLSIIAETAPASLQESRPLITGVHIVLSNKHDFKAVA 179  
DB 121 PDQYPLLPQVRDRAIQLSVKLVKNVIAQTNFVSTSTTRPVLITGVNWLIOEN-ELICTA 179  
QY 180 TDSHRMSORLITLNTSADLMVLPKSLRFSFAVFTDDIEVEVFSPSQILFRSEHIS 239  
DB 180 TDSHRLAVRKQLEVDSENKNVILPGKALAEINKIMSNEEDIDIFFASNQVLFKGVN 239  
QY 240 FYTRLLEGNYPDTRLMTETETEVFNTOSLRHAMERAFNISATONGTVKLEITQNH 299  
DB 240 FISRLLEGHYPDTRLFPENYEIKLSIDNGEYHAI DRASLLAREGGNNVILKSTGDDVV 299  
QY 300 SAHVNSPVEGKVNEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFSPVPTL 359  
DB 300 ELSSTSPBIGTVKEEVDANDVEGSLKISFNSKYMMDALKADNDNEVEVEFFGTMKPFIL 359  
QY 360 TP-GDEESFIQLITPVRT 377  
DB 360 KPKG-D-DSVTQLILPIRT 376

RESULT 25  
ABB81847  
ID ABB81847 standard; protein; 377 AA.  
XX  
XX ABB81847;  
XX  
DT 20-SEP-2002 (first entry)  
XX  
DE Staphylococcus aureus STAAU\_R2.  
XX  
KW STAAU-R2; drug screening; antibacterial; bacterial infection.  
XX  
OS Staphylococcus aureus.  
XX  
FN WO200244718-A2.  
XX

PD 06-JUN-2002.  
XX  
XX 30-NOV-2001; 2001WO-CA001754.  
XX  
PR 30-NOV-2000; 2000US-00727892.  
XX  
PR 20-JUN-2001; 2001US-00885561.  
XX  
PA (PHAG-) PHAGETECH INC.  
XX  
XX Pelletier J, Gros P, Dubow M;  
XX  
XX WPI; 2002-566592/60.  
DR N-PSDB; ABN86787.  
XX  
PT Use of a Staphylococcus aureus (STAAU\_R2) polypeptide or composition  
PT comprising STAAU\_R2 and polypeptide encoded by bacteriophage open reading  
PT frame that specifically interacts with STAAU\_R2, for identifying a  
PT compound active on STAAU\_R2.  
XX  
PS Claim 1; Fig 1; 131pp; English.  
XX  
CC The invention relates to a novel Staphylococcus aureus polypeptide  
CC (STAAU\_R2), and it's use in drug screening assays. The polypeptide of the  
CC invention has antibacterial activity. The compound active on STAAU\_R2 is  
CC useful for the manufacture of an antibacterial agent or for the  
CC manufacture for medication for treating or preventing a bacterial  
CC infection. The sequence represents the S. aureus STAAU\_R2 of the  
CC invention  
XX  
SQ Sequence 377 AA;

Query Match 39.2%; Score 730.5; DB 5; Length 377;  
Best Local Similarity 39.3%; Pred. No. 1e-55;  
Matches 149; Conservative 90; Mismatches 135; Indels 5; Gaps 4;  
QY 1 MIOFSINTFLTHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTSGNGOISIENTIPV 60  
DB 1 MMEFTIKRDYFITQLNDTLKAISSPTTLITGKIDAKEHEVILTGSDSISIEITIPK 60  
QY 61 SNEAGLL-ITSPGAILLEASFFINISSLDPDISINVKIEIQHQVVLTSKSEITLKGK 119  
DB 61 TVDGEDIVNISSETGSVLPGRFFVDIIKKLPGKOVKLTNEQFQTLITSGHSEFNLSGLD 120  
QY 120 VDQYPRQEVSTENPLIKTLKLSIIAETAPASLQESRPLITGVHIVLSNKHDFKAVA 179  
DB 121 PDQYPLLPQVRDRAIQLSVKLVKNVIAQTNFVSTSTTRPVLITGVNWLIOEN-ELICTA 179  
QY 180 TDSHRMSORLITLNTSADLMVLPKSLRFSFAVFTDDIEVEVFSPSQILFRSEHIS 239  
DB 180 TDSHRLAVRKQLEVDSENKNVILPGKALAEINKIMSNEEDIDIFFASNQVLFKGVN 239  
QY 240 FYTRLLEGNYPDTRLMTETETEVFNTOSLRHAMERAFNISATONGTVKLEITQNH 299  
DB 240 FISRLLEGHYPDTRLFPENYEIKLSIDNGEYHAI DRASLLAREGGNNVILKSTGDDVV 299  
QY 300 SAHVNSPVEGKVNEDLDIVSQSGDLTISFNPTYLIESLKAIKSETVKIHFSPVPTL 359  
DB 300 ELSSTSPBIGTVKEEVDANDVEGSLKISFNSKYMMDALKADNDNEVEVEFFGTMKPFIL 359  
QY 360 TP-GDEESFIQLITPVRT 377  
DB 360 KPKG-D-DSVTQLILPIRT 376

RESULT 26  
ABU16160  
ID ABU16160 standard; protein; 377 AA.  
XX  
XX ABU16160;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #1687.



```
Query Match      39.2%; Score 730.5; DB 6; Length 377;
Best Local Similarity 39.3%; Pred. No. 1e-55;
Matches 149; Conservative 90; Mismatches 135; Indels 5; Gaps 4;

QY 1 MIQSFINTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGNGOISIENTIPV 60
DB 1 MMEFTIKRDYFYTQNDLTAKAISPRITLITGKIDAKEHEVILTSGDSEISIEITIPK 60
QY 61 SNEAGLL-ITSPGAILLEASFFINIISLPDISINVKIEIQHGVLTSGKSEITLKGKD 119
DB 61 TVDGEDIVNISSETGSVLPGRFFVDIIKKLPGKOVKLTSTNEQFQTLTSGHSEFNLGLD 120
QY 120 VDQYPRLOEVSTENPLILKTKLLKIIAETAFASLQESRPILTGVHVLNHNKDFKAVA 179
DB 121 PDQYPLLPQVSRDDAIQLSVKLVKNVIAQTFAVSTSTPRVLTGVNWLQEN-ELICTA 179
QY 180 TDSHRMSORLTLTNDNTSADLMVLPKSLRFSFAVFTDDIETVEVFFSPQILFRSEHIS 239
DB 180 TDSHRLAVRKLEQEDVSENKNVITPGKALAEINKIMSDNEEDIDIFFASNOVLFKGVNV 239
QY 180 TDSHRMSORLTLTNDNTSADLMVLPKSLRFSFAVFTDDIETVEVFFSPQILFRSEHIS 239
DB 180 TDSHRLAVRKLEQEDVSENKNVITPGKALAEINKIMSDNEEDIDIFFASNOVLFKGVNV 239
QY 240 FYTRLEGGYPTDRLMTETEFETEVFNTQSLRHAMERAFILSNATQGTVKLEITQNH 299
DB 240 FYTRLEGGYPTDRLMTETEFETEVFNTQSLRHAMERAFILSNATQGTVKLEITQNH 299
QY 240 FISRLLEGHYPTDRLMTETEFETEVFNTQSLRHAMERAFILSNATQGTVKLEITQNH 299
DB 240 FISRLLEGHYPTDRLMTETEFETEVFNTQSLRHAMERAFILSNATQGTVKLEITQNH 299
QY 300 SAHNVSPGVKNEDLDIVSGSDLTISFNPTYLIESLKAISKSTVKIHLFSPVRPTL 359
DB 300 ELSSTSPBIGTVKERVANDVGGSLKISFNSKYNMMDKALINDNEVEVEFFGTMKPFI 359
QY 360 TP-GDEESFTQILTPVRT 377
DB 360 KPKGDD--DSVTQLILPIRT 376

RESULT 28
ADD26272
ID ADD26272 standard; protein; 377 AA.
AC
XX
XX ADD26272;
XX
XX
XX 15-JAN-2004 (first entry)
XX
XX Staphylococcus aureus dnaN protein SEQ ID NO:77.
XX
XX Crystallised recombinant polypeptide; Staphylococcus aureus;
XX Streptococcus pneumoniae; Helicobacter pylori; Pseudomonas aeruginosa;
XX nucleic acid processing; crystal form;
XX ribonuclease diphosphate reductase; nrde;
XX polyribonucleotide phosphorylase; pnpA;
XX transcription termination factor rho;
XX DNA polymerase III; dnaN;
XX ribonucleoside diphosphate reductase minor subunit; uridylylate kinase;
XX orotate phosphoribosyltransferase; pyrE; uracil-DNA-glycosylase;
XX antibacterial; vaccine.
XX
XX Staphylococcus aureus..
XX
XX WO2003025004-A2.
XX
XX 27-MAR-2003.
XX
XX 18-SEP-2002; 2002WO-CA001411.
XX
XX 18-SEP-2001; 2001US-0323040P.
XX 27-SEP-2001; 2001US-032307P.
XX 27-SEP-2001; 2001US-0325421P.
XX 28-SEP-2001; 2001US-0325891P.
XX 01-OCT-2001; 2001US-0326377P.
XX 03-OCT-2001; 2001US-0326774P.
XX 04-OCT-2001; 2001US-0327193P.
XX 30-OCT-2001; 2001US-0340922P.
XX 05-NOV-2001; 2001US-0338709P.

PR 06-NOV-2001; 2001US-0333259P.
PR 18-DEC-2001; 2001US-0341679P.
PA (AFFI-) AFFINIUM PHARM INC.
XX
XX Edwards A, Dharamsi A, Vedadi M, Alam MZ, Arrowemith C, Awrey D;
XX Beattie B, Canadien V, Cox B, Domagala M, Houston S, Li Q;
XX Nethery K, Ng I, Ouyang H, Pinder B, Sheldrick B, Viola C;
XX Wrezel O;
XX WPI; 2003-371793/35.
XX N-PSDB; ADD26271.
XX
XX New crystallized recombinant polypeptides from Staphylococcus aureus,
XX Streptococcus pneumoniae, Helicobacter pylori, Pseudomonas aeruginosa
XX involved in nucleic acid processing, useful as targets for pathogenic
XX bacteria.
XX
XX Claim 218; SEQ ID NO 77; 298pp; English.
XX
XX The present invention describes a crystallised recombinant polypeptide
XX (I) comprising the amino acid sequence of polypeptides from
XX Staphylococcus aureus, Streptococcus pneumoniae, Helicobacter pylori and
XX Pseudomonas aeruginosa which are involved in nucleic acid processing, or
XX amino acid sequences having at least 90 % identity with the polypeptide
XX sequence, where the polypeptide is in crystal form. (I) comprises the
XX amino acid sequence (S) of polypeptides involved in nucleic acid
XX processing, which includes ribonuclease diphosphate reductase, major
XX subunit (nrde), polyribonucleotide phosphorylase (pnpA), transcription
XX termination factor rho (rho), putative polynucleotide
XX nucleotidyltransferase (pnp), DNA ligase (lig), DNA polymerase III, beta-
XX subunit (dnaN) or ribonucleoside diphosphate reductase minor subunit (R2)
XX (nrdf) from S. aureus; uridylylate kinase (pyrH) or orotate
XX phosphoribosyltransferase (pyrE) from H. pylori; uracil-DNA-glycosylase
XX (ung) from P. aeruginosa; and nrde from S. pneumoniae. (I) is in a
XX crystal form. (I) has antibacterial activity, and can be used in
XX vaccines. (I) is useful for designing a modulator for the prevention or
XX treatment of S. aureus, S. pneumoniae, H. pylori and P. aeruginosa
XX related diseases or disorders. The method comprises providing a three-
XX dimensional (3D) structure for (I), identifying a potential modulator by
XX reference to the 3D structure, contacting the potential modulator with
XX the recombinant polypeptide and assaying the activity of the polypeptide
XX or determining the viability of S. aureus, S. pneumoniae, H. pylori or P.
XX aeruginosa after contact with the modulator, where a change in the
XX activity of the polypeptide or the viability of the bacteria indicates
XX that the modulator may be useful for preventing or treating the disease
XX or disorder. The structural and functional information of (I) aid in the
XX discovery and design of therapeutic and diagnostic molecules. The present
XX sequence is used in the exemplification of the present invention.
XX
XX . Sequence 377 AA;
XX
XX Query Match      39.2%; Score 730.5; DB 7; Length 377;
XX Best Local Similarity 39.3%; Pred. No. 1e-55;
XX Matches 149; Conservative 90; Mismatches 135; Indels 5; Gaps 4;
XX
XX QY 1 MIQSFINTLFIHALNTTKRAISTKNAIPILSSIKIEVTSTGVTLTGNGOISIENTIPV 60
XX DB 1 MMEFTIKRDYFYTQNDLTAKAISPRITLITGKIDAKEHEVILTSGDSEISIEITIPK 60
XX
XX QY 61 SNEAGLL-ITSPGAILLEASFFINIISLPDISINVKIEIQHGVLTSGKSEITLKGKD 119
XX DB 61 TVDGEDIVNISSETGSVLPGRFFVDIIKKLPGKOVKLTSTNEQFQTLTSGHSEFNLGLD 120
XX
XX QY 120 VDQYPRLOEVSTENPLILKTKLLKIIAETAFASLQESRPILTGVHVLNHNKDFKAVA 179
XX DB 121 PDQYPLLPQVSRDDAIQLSVKLVKNVIAQTFAVSTSTPRVLTGVNWLQEN-ELICTA 179
XX
XX QY 180 TDSHRMSORLTLTNDNTSADLMVLPKSLRFSFAVFTDDIETVEVFFSPQILFRSEHIS 239
XX DB 180 TDSHRLAVRKLEQEDVSENKNVITPGKALAEINKIMSDNEEDIDIFFASNOVLFKGVNV 239
XX
XX QY 240 FYTRLEGGYPTDRLMTETEFETEVFNTQSLRHAMERAFILSNATQGTVKLEITQNH 299
XX DB 240 FYTRLEGGYPTDRLMTETEFETEVFNTQSLRHAMERAFILSNATQGTVKLEITQNH 299
```

Db 240 FISRLLEGHPDTRLPENYEIKLSIDNGEFVHADRASLLAREGNNVILKSTGDDVV 299  
 Qy 300 SAHVNSEVGVKNEDDLVSQSGDLTISNPYLTSLKAIKSETVKHFLSFVRPFTL 359  
 Db 300 ELSTSPSPEIGTVEVDANDVEGSLKISFNKYMMDALKADNDDEVEVEFFGTMPFIL 359  
 Qy 360 TP-GDEESFIQLITPVRT 377  
 Db 360 KPKGDD--DSVTQLILPRT 376

## RESULT 29

ADD26274  
 ID ADD26274 standard; protein; 377 AA.

XX AC ADD26274;

XX DT 15-JAN-2004 (first entry)

XX Staphylococcus aureus dnaN protein SEQ ID NO.79.

XX crystallised recombinant polypeptide; Staphylococcus aureus;  
 KW Streptococcus pneumoniae; Helicobacter pylori; Pseudomonas aeruginosa;  
 KW nucleic acid processing; crystal form;  
 KW ribonuclease diphosphate reductase; nrdE;  
 KW polynucleotide phosphorylase; pnpA;  
 KW transcription termination factor Rho;  
 KW putative polynucleotide nucleotidyltransferase; DNA ligase;  
 KW DNA polymerase III; dnaN;  
 KW ribonuclease diphosphate reductase minor subunit; uridylyl kinase;  
 KW orotate phosphoribosyltransferase; pyrE; uracil-DNA-glycosylase;  
 KW antibacterial; vaccine.

XX Staphylococcus aureus.

XX WO2003025004-A2.

XX 27-MAR-2003.

XX 18-SEP-2002; 2002WO-CA001411.

XX 18-SEP-2001; 2001US-0323040P.

XX 27-SEP-2001; 2001US-0325307P.

XX 27-SEP-2001; 2001US-0325421P.

XX 28-SEP-2001; 2001US-0325891P.

XX 01-OCT-2001; 2001US-0326337P.

XX 03-OCT-2001; 2001US-0326774P.

XX 04-OCT-2001; 2001US-0327193P.

XX 30-OCT-2001; 2001US-0340922P.

XX 05-NOV-2001; 2001US-0338709P.

XX 06-NOV-2001; 2001US-0333269P.

XX 18-DEC-2001; 2001US-0341679P.

XX (AFFI-) AFFINIUM PHARM INC.

XX Edwards A, Dharamsi A, Vedadi M, Alam MZ, Arrowsmith C, Awrey D;  
 PI Beattie B, Canadian V, Cox B, Domagaia M, Houston S, Li Q;  
 PI Netherly K, Ng I, Ouyang H, Pinder B, Sheldrick B, Viola C;  
 PI Wrezel O;

XX WPI; 2003-371793/35.

XX N-PSDB; ADD26273.

XX New crystallized recombinant polypeptides from Staphylococcus aureus,  
 PT Streptococcus pneumoniae, Helicobacter pylori, Pseudomonas aeruginosa  
 PT involved in nucleic acid processing, useful as targets for pathogenic  
 PT bacteria.

XX Claim 218; SEQ ID NO 79; 298pp; English.

XX The present invention describes a crystallised recombinant polypeptide  
 CC (I) comprising the amino acid sequence of polypeptides from

CC Staphylococcus aureus, Streptococcus pneumoniae, Helicobacter pylori and  
 CC Pseudomonas aeruginosa which are involved in nucleic acid processing, or  
 CC amino acid sequences having at least 90 % identity with the polypeptide  
 CC sequence, where the polypeptide is in crystal form. (I) comprises the  
 CC amino acid sequence (S) of polypeptides involved in nucleic acid  
 CC processing, which includes ribonuclease diphosphate reductase, major  
 CC subunit (nrdE), polynucleotide phosphorylase (pnpA), transcription  
 CC termination factor Rho (Rho), putative polynucleotide  
 CC nucleotidyltransferase (pnp), DNA ligase (lig), DNA polymerase III, beta-  
 CC subunit (dnaN) or ribonucleoside diphosphate reductase minor subunit (R2)  
 CC (nrdF) from S. aureus; uridylyl kinase (pyrE) or orotate  
 CC phosphoribosyltransferase (pyrE) from H. pylori; uracil-DNA-glycosylase  
 CC (ung) from P. aeruginosa; and nrdE from S. pneumoniae. (I) is in a  
 CC crystal form. (I) has antibacterial activity, and can be used in  
 CC vaccines. (I) is useful for designing a modulator for the prevention or  
 CC treatment of S. aureus, S. pneumoniae, H. pylori and P. aeruginosa  
 CC related diseases or disorders. The method comprises providing a three-  
 CC dimensional (3D) structure for (I), identifying a potential modulator by  
 CC reference to the 3D structure, contacting the potential modulator with  
 CC the recombinant polypeptide and assaying the activity of the polypeptide  
 CC or determining the viability of S. aureus, S. pneumoniae, H. pylori or P.  
 CC aeruginosa after contact with the modulator, where a change in the  
 CC activity of the polypeptide or the viability of the bacteria indicates  
 CC that the modulator may be useful for preventing or treating the disease  
 CC or disorder. The structural and functional information of (I) aid in the  
 CC discovery and design of therapeutic and diagnostic molecules. The present  
 CC sequence is used in the exemplification of the present invention.

XX Sequence 377 AA;

Query Match 39.2%; Score 730.5; DB 7; Length 377;

Best Local Similarity 39.3%; Pred. No. 1e-55;

Matches 149; Conservative 90; Mismatches 135; Indels 5; Gaps 4;

Qy 1 MIQPSINRTLFIHALNTTKAISTKNAIPILSSIKIEVSTGVLTSNGSQISIENTIPV 60

Db 1 MMEFTIKRDYFITQANDTLKAIPTPTLTILTKIDAKEHEVLTGSDSEISIEITPK 60

Qy 61 SNENAGLL-ITSPGAILLEASFFINIISSLPDISINVKIEHQHVLTSGKSEITLKGKD 119

Db 61 TVGEDIINISGTVGVLPGRFFVDIIKLPKGVLTSTNEQFTLTITSGHSEFNLSGLD 120

Qy 120 VDQYPRLEQVSTENPLILTKLLKSIATAFAASLQESRPILTGVHIVLSNKHDKFAVA 179

Db 121 PDQYPLLPQVSRDDAIQLSVKVLKNVIAQINFAVSTETPVLTVGNWLQEN-ELICTA 179

Qy 180 TDSHRSQRLITLNTSADLMVLPKSLREFSAVTDITVEVTFSPSQILFRSEHIS 239

Db 180 TDSHRLAVRKLQLEDVSENKNVPIPGKALAEKNIMSDNEEDIDIFFASQNVLFKVGNNV 239

Qy 240 FYTRLLEGNYPDTRLIMTEFETEVVFNQSLRHAMERAFILSNATQNGTVKLEITONHI 299

Db 240 FISRLLEGHPDTRLPENYEIKLSIDNGEFVHADRASLLAREGNNVILKSTGDDVV 299

Qy 300 SAHVNSEVGVKNEDDLVSQSGDLTISNPYLTSLKAIKSETVKHFLSFVRPFTL 359

Db 300 ELSTSPSPEIGTVEVDANDVEGSLKISFNKYMMDALKADNDDEVEVEFFGTMPFIL 359

Qy 360 TP-GDEESFIQLITPVRT 377

Db 360 KPKGDD--DSVTQLILPRT 376

## RESULT 30

ABW01648

ID ABW01648 standard; protein; 377 AA.

XX AC ABW01648;

XX DT 12-FEB-2004 (first entry)

XX Staphylococcus aureus dnaN gene beta subunit protein.









[illegible]

```
XX G. stearothermophilus alpha subunit.
DE DNA polymerase; DNA sequencing; DNA amplification.
KW Geobacillus stearothermophilus.
OS US2004038289-A1.
PN 26-FEB-2004.
XX 25-SEP-2003; 2003US-00671403.
XX 08-APR-1997; 97US-0043202P.
PR 08-APR-1998; 98US-00057416.
PR 18-AUG-2000; 2000US-00642218.
PR 21-NOV-2000; 2000US-00716964.
XX (ODON/) O'DONNELL M E.
PA (YUZH/) YUZHAKOV A.
PA (YURI/) YURIEVA O.
PA (JERU/) JERUZALMI D.
PA (BRUC/) BRUCK I.
PA (KURI/) KURIYAN J.
XX O'donnell ME, Yuzhakov A, Yurieva O, Jeruzalmi D, Bruck I;
PI Kuriyan J;
XX WPI; 2004-203218/19.
DR N-PSDB; ADJ67943.
XX Isolated DNA molecule from Bacillus stearothermophilus encoding a delta
PT subunit of a DNA polymerase I mul-type, useful for producing DNA
PT polymerases for use in DNA sequencing and DNA amplification methods.
XX Disclosure; SEQ ID NO 174; 245pp; English.
XX The invention relates to an isolated DNA molecule from Bacillus
CC stearothermophilus encoding a delta subunit of a DNA polymerase I. The
CC subunits are useful for producing DNA polymerases for use in DNA
CC sequencing and DNA amplification methods. The present sequence is used in
CC the exemplification of the present invention.
XX Sequence 334 AA;
Query Match 31.5%; Score 586; DB 8; Length 334;
Best Local Similarity 37.9%; Pred. No. 5.2e-43;
Matches 127; Conservative 74; Mismatches 126; Indels 8; Gaps 7;
QY 49 NQGIS-IENTIPVSNENAGLL-ITSPGAILLEASFFINIISLSPDISINVKIEHQVVL 106
DB 1 NSDISIIESFIPLEKEGKLLVDVKRPGSIVLQARFSEIVKLPQQTVEIETEDNFLTII 60
QY 107 TSGKSEITLKGVDQYPRLOEVSTENPLILKTKLKSIIAETAFAASIQESRPILTVGH 166
DB 61 RSGHSEFRGLNGNADEYPRLPQIEENVFQIPADLLKTVIRQTVFAVSTSETRPILTVGN 120
QY 167 IVLSNHKDFKAVATDSHRMSQRLITLDNTSADLM--VVLPSKSLREFSAVFTDD-IETVE 223
DB 121 WKV-EGHELVCVTATDSHRLAMRKVKIIESENEVSNNVPIPGKSLNELSKIILDDGNHPVD 179
QY 224 VFFSPSQILFRSEHISFYTRLLEGNYPDTRLMTTEFEFEVFNFTQSLRHAMERAFILSN 283
DB 180 IVMATANQVLFKAHLLFFSRLLDGNYPETARLIPTESKTTMIVNAKEFLQADIRASLLAR 239
QY 284 ATQNGTVKL-EITQNHISAHVNSPEVKVNEIDLIVSQSGDLITFSFNPTYLIESKAIK 342
DB 240 EGRNNVVKLTTPPGGMLTSSISPEIGKVTEQLQTESLEGEELNIFSASKYMMDALRALD 299
QY 343 SETVKIHLSPVRPFTLTGDEESFIQLITPVRT 377
DB 300 GTDIQISFTGMRPFLRLP-LHTDSMLQLILPVRT 333
```

```
RESULT 37
ADJ68156
ID ADJ68156 standard; protein; 334 AA.
XX
AC ADJ68156;
XX
DT 20-MAY-2004 (first entry)
DE G. stearothermophilus alpha subunit.
XX
KW DNA polymerase; DNA sequencing; DNA amplification.
XX
OS Geobacillus stearothermophilus.
XX
PN US2004038290-A1.
XX
PD 26-FEB-2004.
XX
PF 25-SEP-2003; 2003US-00671419.
XX
PR 08-APR-1997; 97US-0043202P.
PR 08-APR-1998; 98US-00057416.
PR 18-AUG-2000; 2000US-00642218.
PR 21-NOV-2000; 2000US-00716964.
XX
PA (ODON/) O'DONNELL M E.
PA (YUZH/) YUZHAKOV A.
PA (YURI/) YURIEVA O.
PA (JERU/) JERUZALMI D.
PA (BRUC/) BRUCK I.
PA (KURI/) KURIYAN J.
XX
PI O'donnell ME, Yuzhakov A, Yurieva O, Jeruzalmi D, Bruck I;
PI Kuriyan J;
XX
DR WPI; 2004-203219/19.
DR N-PSDB; ADJ68155.
XX
PT Novel DNA molecule from Thermotoga species encoding delta prime subunit
PT of DNA polymerase III-type enzyme, useful for producing the enzyme by
PT recombinant techniques.
XX
PS Disclosure; SEQ ID NO 174; 245pp; English.
XX
CC The invention relates to an isolated DNA molecule from Bacillus
CC stearothermophilus encoding a delta subunit of a DNA polymerase I. The
CC subunits are useful for producing DNA polymerases for use in DNA
CC sequencing and DNA amplification methods. The present sequence is used in
CC the exemplification of the present invention.
XX
SQ Sequence 334 AA;
Query Match 31.5%; Score 586; DB 8; Length 334;
Best Local Similarity 37.9%; Pred. No. 5.2e-43;
Matches 127; Conservative 74; Mismatches 126; Indels 8; Gaps 7;
QY 49 NQGIS-IENTIPVSNENAGLL-ITSPGAILLEASFFINIISLSPDISINVKIEHQVVL 106
DB 1 NSDISIIESFIPLEKEGKLLVDVKRPGSIVLQARFSEIVKLPQQTVEIETEDNFLTII 60
QY 107 TSGKSEITLKGVDQYPRLOEVSTENPLILKTKLKSIIAETAFAASIQESRPILTVGH 166
DB 61 RSGHSEFRGLNGNADEYPRLPQIEENVFQIPADLLKTVIRQTVFAVSTSETRPILTVGN 120
QY 167 IVLSNHKDFKAVATDSHRMSQRLITLDNTSADLM--VVLPSKSLREFSAVFTDD-IETVE 223
DB 121 WKV-EGHELVCVTATDSHRLAMRKVKIIESENEVSNNVPIPGKSLNELSKIILDDGNHPVD 179
QY 224 VFFSPSQILFRSEHISFYTRLLEGNYPDTRLMTTEFEFEVFNFTQSLRHAMERAFILSN 283
DB 180 IVMATANQVLFKAHLLFFSRLLDGNYPETARLIPTESKTTMIVNAKEFLQADIRASLLAR 239
```

Db	61	RSCHSEFRLNGLADEYPRLPQIEENYFQIPADLLKTVIRQTVFAVSTSETRPILTVGN	120
QY	167	IVLSNHKDFKAVATDSHRMSQRLITLDNTSADLM--VVLPSKSLREFSAVFTDD-IETVE	223
Db	121	WKV-EHGLVCTCTATDSHRLAMRKVKIIIESENEVSYNVVPGKSLNELSKIILDDGNHPVD	179
QY	224	VFFSPQLLFRSEHISFYTRLEGNYPOTDRLLMTFETFEVVFNTQSLRHAMERAPLISN	283
Db	180	IVMTANQVLFKAHLLFFSRLLDGNYPETARLIPTESKTTMIVNAKEFLQAIIDRASLLAR	239
QY	284	ATQNGTVKL-EITQNHISAHVNSPEVGVKNEDLDIVSQSGDLTISFNPTYLIESIKAIK	342
Db	240	EGRNVVVKLTLPGLGMLIEISSIEIGKVTQEQTESLEGELNLSFSKAKYMDALD	299
QY	343	SETVKIHLPSVRPPTLTPGDEESFQILTPVRT	377
Db	300	GTDIQISFTGMRPFLLRP-LHTDSMLQLILPVRT	333
RESULT	39		
ADJ79453	ID	ADJ79453 standard; protein; 334 AA.	
XX	AC	ADJ79453;	
XX	DT	20-MAY-2004 (first entry)	
DE		G. steothermophilus alpha subunit.	
XX	KW	DNA polymerase; DNA sequencing; DNA amplification.	
XX	OS	Geobacillus steothermophilus.	
XX	FN	US2004043414-A1.	
XX	PD	04-MAR-2004.	
XX	PF	25-SEP-2003; 2003US-00670844.	
XX	PR	08-APR-1997; 97US-0043202P.	
PR	PR	08-APR-1998; 98US-00057416.	
PR	PR	18-AUG-2000; 2000US-00642218.	
XX	PR	21-NOV-2000; 2000US-00716964.	
XX	PA	(ODON/) O'DONNELL M E.	
PA	PA	(YUZH/) YUZHAKOV A.	
PA	PA	(YURI/) YURIEVA O.	
PA	PA	(JERU/) JERUZALMI D.	
PA	PA	(BRUC/) BRUCK I.	
XX	PA	(KURI/) KURIYAN J.	
XX	PI	O'donnell ME, Yuzhakov A, Yurieva O, Jeruzalmi D, Bruck I;	
PI	PI	Kuriyan J;	
XX	XX	WPI; 2004-225698/21.	
DR	DR	N-PSDB; ADJ79452.	
XX	XX		
PT	PT	Novel isolated DNA molecule from Bacillus steothermophilus, encoding	
PT	PT	tau subunit of DNA polymerase III-type enzyme, useful in amplification	
PT	PT	and sequencing reactions.	
XX	XX		
PS	PS	Disclosure; SEQ ID NO 174; 245pp; English.	
XX	XX		
CC	CC	The invention relates to an isolated DNA molecule from Bacillus	
CC	CC	steothermophilus encoding a delta subunit of a DNA polymerase I. The	
CC	CC	subunits are useful for producing DNA polymerases for use in DNA	
CC	CC	sequencing and DNA amplification methods. The present sequence is used in	
CC	CC	the exemplification of the present invention.	
XX	XX		
SQ	SQ	Sequence 334 AA;	
Query	Match	31.5%; Score 586; DB 8; Length 334;	
Best	Local Similarity	37.9%; Pred. No. 5.2e-43;	



